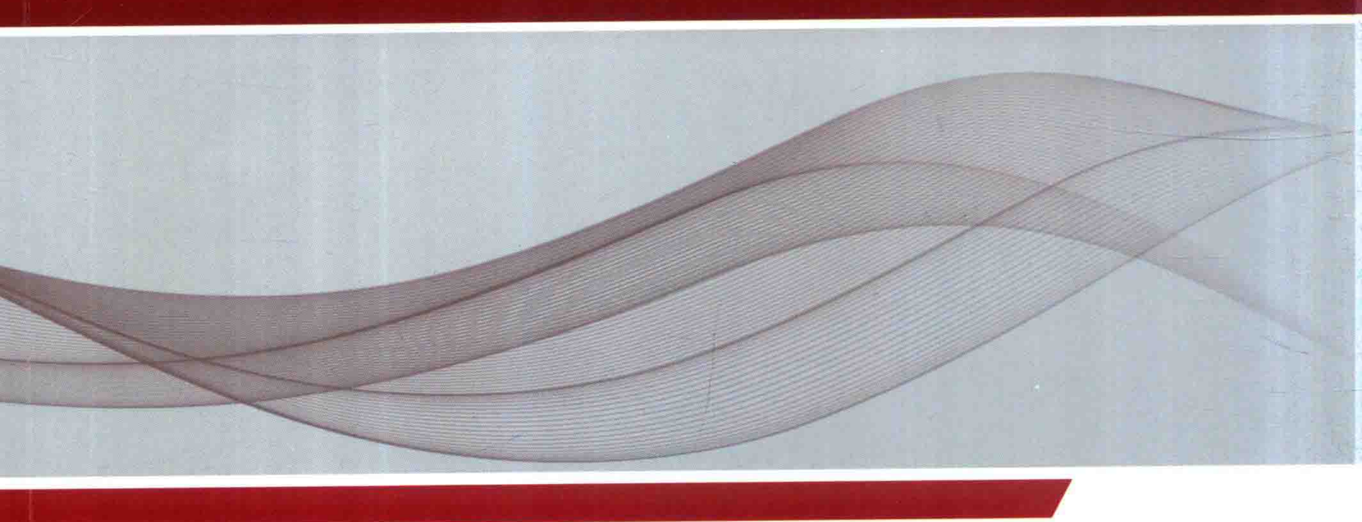


"十三五"
国家重点图书出版规划项目
ICT认证系列丛书



华为技术认证

华为MPLS技术学习指南

王达 主编




中国工信出版集团



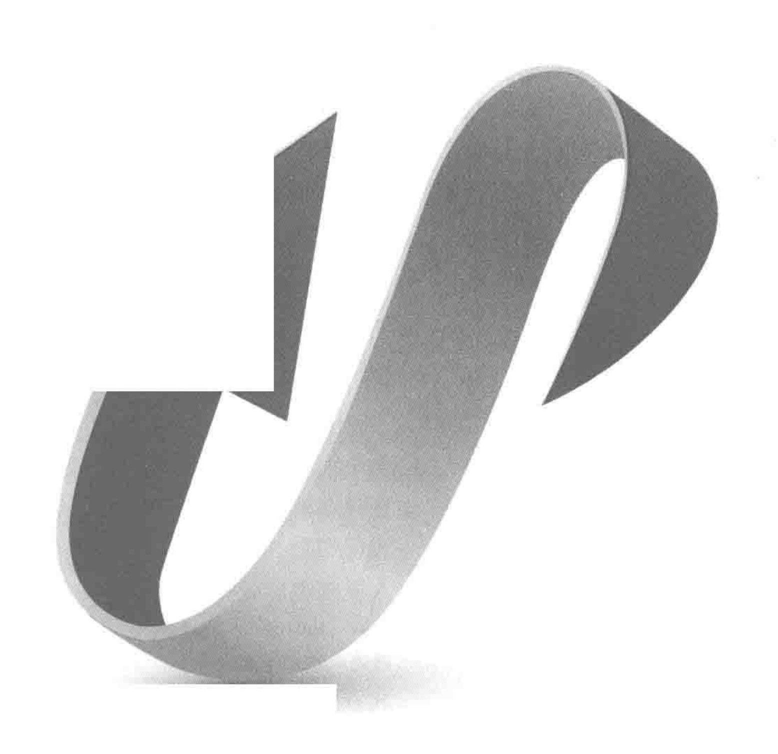


“十三五”
国家重点图书出版规划项目
ICT认证系列丛书

华为技术认证

华为MPLS技术学习指南

王 达 主编



人 民 邮 电 出 版 社
北 京

图书在版编目（CIP）数据

华为MPLS技术学习指南 / 王达主编. -- 北京 : 人民邮电出版社, 2017.12
（ICT认证系列丛书）
ISBN 978-7-115-45648-9

Ⅰ. ①华… Ⅱ. ①王… Ⅲ. ①宽带通信系统一综合业务通信网一指南 Ⅳ. ①TN915.142-62

中国版本图书馆CIP数据核字(2017)第267815号

内容提要

本书由华为技术有限公司授权编写并出版，是一本系统、深入介绍华为设备MPLS隧道技术的工具图书。本书内容主要包括MPLS隧道基础知识、静态LSP、动态LDP LSP、MPLS TE和MPLS DS-TE隧道建立，以及MPLS QoS等方面，同时本书也是华为技术有限公司指定的ICT认证培训教材。

全书共9章，第1章介绍了MPLS技术的基础知识和相关技术原理；第2章介绍了静态LSP建立的配置与管理方法；第3～4章介绍了动态LDP LSP建立的相关技术原理，以及LDP LSP基本功能和各种扩展功能的配置与管理方法；第5～7章介绍了MPLS TE隧道相关技术原理及各项功能配置与管理方法；第8章介绍了华为S系列交换机中的MPLS QoS功能技术原理及配置与管理方法；第9章介绍了华为AR G3系列中的MPLS DS-TE隧道相关技术原理及配置与管理方法。

为了帮助大家理解，书中拓展介绍了许多相关的计算机网络通信原理，并且各章均有大量的典型配置案例，并对一些典型故障排除方法进行了详细的介绍。另外，本书经过华为技术有限公司多位专家指导和审核，无论是在专业性方面，还是在经验性和实用性方面均有很好的保障，是相关人员自学或者教学华为设备MPLS配置与管理的必选教材。

◆ 主　　编　王　达
　责任编辑　李　静　王建军
　责任印制　彭志环

◆ 人民邮电出版社出版发行　　北京市丰台区成寿寺路11号
　邮编　100164　　电子邮件　315@ptpress.com.cn
　网址　http://www.ptpress.com.cn
　北京隆昌伟业印刷有限公司印刷

◆ 开本：787×1092　1/16
　印张：28.5　　　　　　2017年12月第1版
　字数：668千字　　　　2017年12月北京第1次印刷

定价：95.00元

读者服务热线：(010)81055488　印装质量热线：(010)81055316
反盗版热线：(010)81055315

序

人类社会和人类文明发展的历史也是一部科学技术发展的历史。半个多世纪以来，精彩纷呈的ICT技术，汇聚成了波澜壮阔的互联网，突破了时间和空间的限制，把人类社会和人类文明带入到前所未有的高度。今天，人类社会已经步入网络和信息时代，我们已经处在无处不在的网络连接中。连接已经成为一种常态，信息浪潮迅速而深刻地改变着我们的工作和生活。人们与世界连接得如此紧密，实现了随时随地自由沟通，对信息与数据的获取和分享也无处不在。这意味着，这个连接的世界，正以超乎想象的速度与力量，对人类社会的政治、经济、商业文明和生产方式等进行全面的重塑。

ICT 正在蓬勃发展，移动化、物联网、云计算和大数据等新趋势正在引领行业开拓新的格局。世界正在发生影响深远的数字化变革，互联网正在促进传统产业的升级和重构。以业务、用户和体验为中心的敏捷网络架构将深刻影响未来数字社会的发展。我们深知每个人都拥有平等的数字发展机会，这对于构建一个更加公平的现实世界至关重要。

ICT 产业的发展离不开人才的支撑，产业的变革也将为 ICT 行业人才的知识体系和综合技能带来更高的挑战。华为作为全球领先的信息与通信解决方案供应商，其产品与解决方案已广泛应用于金融、能源、交通、政府、制造等各个行业。同时，其非常注重对 ICT 专业人才的培养。所以，华为与行业专家、高校老师合作编写了《华为 ICT 认证系列丛书》，旨在为广大用户、ICT 从业者，以及愿意投身到 ICT 行业中的人士提供更加便利的学习帮助。

继 2014 年华为与王达老师合作并出版《华为交换机学习指南》《华为路由器学习指南》《华为 VPN 学习指南》以来，得到了广大读者的高度肯定和大力支持。随着读者朋友的成长，大家渴望更加专业的技术学习，其中最受关注的是 MPLS 技术。为此，华为再度与王达老师合作并出版《华为 MPLS 技术学习指南》和《华为 MPLS VPN 学习指南》两本图书。这两本图书从学习和实用的角度，基于学习的逻辑对知识点进行了系统的组织编排，内容由浅入深，让读者逐步掌握各种 MPLS 技术原理，以及在 L2VPN、L3VPN、MPLS TE、MPLS DS-TE 和 QoS 应用方面的配置与管理方法。这两本图书还配备了大量不同场景下的各种 MPLS 应用方案的配置示例及典型故障排除方法，让读者能够真正学以致用。希望本书能够帮助读者快速地学习华为设备的 MPLS 技术，不断提升，在 ICT 行业大展身手！

自　序

本书与配套的《华为 MPLS VPN 学习指南》同时创作并完成，建议先学习本书。本书全面、深入地介绍了各种 MPLS 隧道建立，以及 MPLS QoS 方面的技术原理及相关功能的配置与管理方法。

本书出版背景

本书出版的原始动力主要来自于读者的需求。笔者自 2014 年出版了《华为交换机学习指南》和《华为路由器学习指南》两本图书后，经常有读者问我是否打算出版其他方面的华为图书，其中最受关注的就是华为 VPN、MPLS、WLAN 等方面相对高端的 HCIE 技术图书。因为技术图书有一个特点，即越是高端的技术，相关图书越少，网上的专业资源也越少。但随着读者朋友的成长，他们越来越渴望学到更高级的技能，因此在 2017 年年初萌生了出版《华为 VPN 学习指南》《华为 MPLS 技术学习指南》和《华为 MPLS VPN 学习指南》这三本图书的想法。在得到华为技术有限公司和人民邮电出版社的认可和支持后笔者就开始了新的创作征程。《华为 VPN 学习指南》一书已于 2017 年 9 月正式出版上市。

经过近四年的时间以及几十位国内通信领域专家学者和华为技术有限公司各级领导的共同努力，截至目前，华为 ICT 认证系列图书已出版了十余部，HCNA 和 HCNP 级别的培训图书基本上已成体系，目前主要缺少的是 HCIE 级别的培训图书。已出版的这十余部图书，经过几年的市场检验，得到了读者朋友们广泛的认可和赞赏。为了完善培训图书体系，帮助广大用户掌握各高级设备功能应用，也有必要继续编写 HCIE 级别的图书。十分荣幸，也非常感谢华为技术有限公司的信任，再次把 HCIE 培训图书的开篇创作任务交给了笔者。

本书与笔者前面出版的几本华为图书一样，也得到了华为技术有限公司许多产品专家的严格审核和技术把关，他们提供了许多宝贵的技术指导和修订意见。本书还得到了人民邮电出版社许多编辑老师的多次编辑、审核，所以，无论是专业性、实用性，还是图书编排、出版质量上，本书都有着非一般图书可比的全线保障，敬请读者放心选购。希望这两本书能继续得到大家的喜爱，更希望这两本书能给大家带来实在的帮助。同时也衷心地感谢华为技术有限公司和人民邮电出版社多位领导的大力支持，感谢参与本书编审的技术专家和编辑老师们的辛勤付出，您们辛苦了！

服务与支持

为了加强与读者朋友们的交流与沟通，同时也方便读者朋友们相互交流与学习，及时了解图书配套视频课程和在线培训资讯，笔者向大家提供了全方位的交流平台。

- 超级读者、学员交流 QQ 群

读者交流 QQ 群：516844263

视频课程学员 QQ 群：398772643

■ 两个专家博客

51CTO 博客：http://winda.blog.51cto.com

CSDN 博客：http://blog.csdn.net/lycb_gz

■ 两个认证微博

新浪微博：weibo.com/winda

腾讯微博：t.qq.com/winda2010

■ 两个视频课程中心

CSDN 学院课程中心：http://edu.csdn.net/lecturer/74

51CTO 学院课程中心：http://edu.51cto.com/lecturer/user_id-55153.html

■ 微信及公众号

微信：windanet （加入后可进入读者微信群）

微信公众号：windanetclass

鸣谢

本书由王达主编并统稿，经过数十位编委、技术专家数月夜以继日的工作，一次次严格的审校、修改和完善，本书终于完成，并顺利高质量地出版上市。在此感谢华为技术有限公司各位专家缜密的技术审校和大力支持，感谢人民邮电出版社各位编辑老师，以及各位编委的辛勤工作！以下是参与本书编写和技术审校人员名单（排名不分先后）。

编委人员：何艳辉、周健辉、何江林、卢翠环、王传寿、谭文凤、李峰、郑小建、余志坚、曾育文、刘云根、谢桂安、罗广平、朱碧霞、胡海侨、黄丽君、王爽、陈玉生、蔡学军、李想、夏强、刘胜华、罗巧芬

技术审校：蓝鹏、史晓健、管超、江永红

前　言

每部图书的创作都是一次艰难的历程，都是一次严峻的挑战。HCIE 级别图书的创作难度要远大于之前 HCNA 和 HCNP 图书的创作难度，因为 MPLS 历来是数据通信方面最复杂、最难懂的技术领域，其中涉及许许多多深奥且复杂的技术原理。

本书特色

■ 华为官方授权、审核

华为技术有限公司官方直接授权创作本书，并对整个图书创作、出版的各个阶段进行跟踪、审核，所以本书无论在图书质量和内容专业性方面均有很好的保障。这也是本书能作为华为 ICT 认证培训图书之一的前提与基础。

■ 系统、深入、不泛泛而谈

这是笔者一直坚持的著书特色，也长期得到了广大读者的认可。本书对华为设备 MPLS 隧道本身相关的主要技术原理和功能应用配置与管理方法都做了系统而又深入的介绍，真正可以做到“一册在手，别无所求”。

■ 细致、通俗且富有经验性

本书所涉及的技术原理比较复杂、难懂，所以笔者在编写本书时充分结合了近二十年专门研究计算机网络通信原理方面的经验，尽可能地从应用的角度把各项技术原理进行细化及通俗的剖析。

■ 细节入微、层次分明、重点突出

在细节方面，笔者尽可能坚持深度剖析，通过各种手段寻找答案，尽量解决读者的困惑。在层次方面，笔者严格遵循渐进式的学习规律，尽可能做到条理清晰、架构明确、没有知识点的跳跃。在重点描述方面，对于需要引起读者格外注意的地方，笔者都会在内容上以加粗方式显示，方便读者把握重点。

■ 典型配置示例和故障排除方法的结合

为了增强本书的实用性，在介绍完每一种相关功能配置后笔者都列举了大量不同场景下的配置示例，以加深大家对前面所学技术原理和具体配置与管理方法的理解。许多配置示例完全可以直接应用于不同的现实场景。另外，为了使大家能在部署 MPLS 方案时迅速地排除所遇到的故障，笔者在大部分章节的最后都介绍了针对经典故障现象的排除方法，使得本书具有非常高的专业性和实用性。

适用读者对象

本书具备极高的系统性、专业性和实用性，适合的读者对象如下：

- 参加华为 R&S HCIE 认证的朋友；
- 希望从零开始系统学习华为设备 MPLS 技术的朋友；

- 华为培训合作伙伴、华为网络学院的学员；
- 使用华为 S 系列交换机、AR 系列路由器产品的用户；
- 高等院校计算机网络专业的学生。

本书主要内容

本书内容非常丰富，共 9 章，对华为设备各种 MPLS 隧道建立和 MPLS QoS 方面的相关技术原理及功能配置与管理方法均进行了详细而又深入的介绍，并在每章给出了大量的典型配置示例。下面是各章的主要内容介绍。

第 1 章 MPLS 基础

本章介绍了 MPLS 技术的基础知识和基本的技术原理，包括 MPLS 的由来、MPLS 体系结构、MPLS 标签、MPLS 报文转发流程、LSP 的连通性检测等，这些都是其他 MPLS 技术原理和应用的基础。

第 2 章 静态 LSP 配置与管理

本章介绍了最简单的静态 LSP 的配置与管理方法，以及静态 LSP 建立的典型故障排除方法。

第 3 章 MPLS LDP 基本功能配置与管理

本章介绍了采用 LDP 作为信令协议而动态建立的 LDP LSP 的配置与管理和典型故障排除方法。其中重点介绍了 LDP 的各方面基础和技术原理，包括 LDP 会话消息、LDP 会话的建立、LDP 的标签发布和管理、动态 LDP LSP 的建立等。

第 4 章 MPLS LDP 扩展功能配置与管理

本章介绍了在 LDP LSP 建立中所涉及的可选扩展功能的配置与管理方法，包括 LDP LSP 的 BFD 检测、LDP 与路由联动、LDP FRR、LDP GR 和 LDP LSP 安全机制等功能的配置与管理方法。

第 5 章 MPLS TE 基本功能配置与管理

本章介绍了 MPLS TE（流量工程）隧道中 CR-LSP 建立相关的技术原理和静态、动态 MPLS TE 隧道的配置与管理方法，以及 MPLS TE 隧道维护方法。在基础知识和技术原理方面重点介绍了 RSVP-TE 的消息类型、对象类型、消息格式、TE 隧道属性/链路属性、OSPF TE/IS-IS TE 信息发布原理和 CR-LSP 建立流程等。

第 6 章 MPLS TE 参数调整配置与管理

本章介绍了 MPLS TE 隧道建立过程中的参数调整功能的配置与管理方法。这里所调整的参数是指用于建立 CR-LSP 的参数，主要包括 RSVP-TE 信令参数、CR-LSP 路径选择参数、MPLS TE 隧道建立参数 3 个方面。

第 7 章 MPLS TE 可靠性功能配置与管理

本章介绍了 MPLS TE 隧道的可靠性功能配置与管理方法，主要包括常用的 CR-LSP 备份及 BFD for CR-LSP 两个方面。

第 8 章 MPLS QoS 配置与管理

本章介绍了华为 S 系列交换机所支持的 MPLS QoS 技术原理和相关功能配置与管理方法。

第 9 章 MPLS DS-TE 配置与管理

本节介绍了华为 AR G3 系列路由器所支持的 MPLS QoS 功能——DS-TE 隧道的相关技术原理、静态/动态 DS-TE 隧道配置与管理方法。本章的技术原理比较复杂，涉及 MPLS QoS 中一系列复杂的技术，包括 Diffserv 服务模型、DSCP/802.1P/LP/EXP 优先级，及其与队列、PHB 之间的相互映射关系，以及 DS-TE 中的 LSP 抢占、TE-Class 映射、带宽约束模型、DS-TE 模式等。

阅读注意地方

在阅读本书时，请注意以下几点。

- 在学习华为 MPLS 时，建议先学习本书，然后学习配套的《华为 MPLS VPN 学习指南》一书。
- 书中以 V200R010 及以上版本华为 S 系列交换机、V200R008 及以上版本 AR G3 系列路由器为主线进行介绍的。
- 在配置命令代码介绍中，粗体字部分是命令本身或关键字选项部分，是不可变的；斜体字部分是命令或者关键字的参数部分，是可变的。
- 在介绍各种 VPN 技术及功能配置说明过程中，对于需要特别注意的地方均以粗体字格式加以强调，以便读者在阅读学习时特别注意。

目　录

第1章 MPLS基础

MPLS 其实可以理解为一种把三层路由信息映射成二层交换路径的交换方式。因为在 MPLS 交换中，首先是基于 IP 路由表在通信路径上建立用于指导 MPLS 报文转发的 LSP（标签转发路径，或称 MPLS 隧道），然后 MPLS 网络中各设备就可以直接利用在 LSP 中为各设备分配的 MPLS 标签进行 MPLS 报文转发。正因如此，MPLS 是一种标签交换技术。

MPLS 最初的设计思想来源于二层的 FR（帧中继）中的 DLCI（数据链路连接标识），以及 ATM（异步传输模式）网络中的 VPI/VCI（虚拟路径标识符/虚拟通道标识符），其都可以看作是一种标签，而且都属于数字类型。MPLS 的标签也是数字类型，在原理上也与 FR 的 DLCI、ATM 的 VPI/VCI 非常类似。

第 1 章作为本书的开篇，主要介绍 MPLS 技术的一些基础知识和基本工作原理，包括 MPLS 的起源、MPLS 基本架构、MPLS 标签、MPLS LSP 的建立、MPLS 报文转发流程等。

1.1 MPLS 基础

MPLS（Multiprotocol Label Switching，多协议标签交换）是一种应用于运营商 IP 骨干网的数据交换技术。其在无连接的 IP 网络上引入面向连接（即邻居设备间必须先建立某种连接）的标签交换概念，将第三层路由技术和第二层交换技术相结合，充分发挥了 IP 路由的灵活性和二层交换的简捷性。

MPLS 起源于 IPv4 网络，但目前其核心技术可通过扩展支持多种网络层协议，如 IPv6、IPX（Internet Packet Exchange，因特网包交换）和 CLNP（Connectionless Network Protocol，无连接网络协议）等，在数据链路层上支持以太网、PPP、HDLC 等多种协议，这也就是其名称中“多协议”的含义。

1.1.1 MPLS 的起源

20 世纪 90 年代中期，路由器技术的发展远远滞后于网络的发展速度与规模，主要表现在转发效率低下及无法提供 QoS 保证。其本质原因就是：当时路由查找算法使用最长匹配原则，必须使用软件查找最佳的路由表项。

IP 路由转发的依据就是 IP 路由表项，而 IP 报文的报头中仅含有“目的 IP 地址”字段，而没有对应的“子网掩码”字段，所以从 IP 报文不可直接确定所用的 IP 路由表项，IP 报文的转发需要从当前 IP 路由表中选择一条最佳的转发路径。这就是 IP 路由中的“最长匹配原则”，即根据 IP 报文中的“目的 IP 地址”选择一个可以匹配，且子网掩码最长（代表最精确）的路由表项来为该 IP 报文进行转发。选择了 IP 路由表项才能确定从本地设备转发的出接口和下一跳 IP 地址，这样的 IP 路由表项选择是每经过一跳设备都要进行的，所以 IP 路由转发方式比较消耗资源。加之当初并没有像 ASIC 这样的集成电路技术，IP 路由表项的选择纯粹依据软件系统计算完成，效率比较低。

正因为 IP 路由转发效率比较低，有人在想是不是可以采用另外一种标识来实现同样的目的，且不用每一跳都经过复杂的计算就可以确定正确的数据转发路径，以提高转发

效率。ATM（Asynchronous Transfer Mode，异步传输模式）技术就是其中的杰出代表。ATM 采用定长标签（即信元），并且只需要维护比路由表规模小得多的标签表，就能够提供比 IP 路由方式高得多的转发性能。也正因如此，当时还出现过 IP 与 ATM 竞争的场面。然而 ATM 过于复杂的设计导致没有多少厂商能够完全领会并成功生产所需的软、硬件产品，而且其无法与 IP 网络很好地融合导致最终没有广泛应用。但 ATM 仍有可取之处：它首先摒弃了繁琐的路由表查找过程，改为简单快速的标签交换；其次把具有全局意义的路由表改为只有本地意义的标签表。

传统的 IP 技术简单，且部署成本低。如何结合 IP 与 ATM 的优点成为当时的热门话题。MPLS（多协议标签交换）技术就是在这种背景下产生的。MPLS 最初的引入背景其实就是为了解决当初仅能依靠软件系统（受制于当时路由器设备的硬件技术落后）进行路由转发而带来的数据转发效率低下的问题。因为，MPLS 与传统 IP 路由方式相比，在数据转发时，MPLS 只需在网络边缘分析 IP 报头，而不用在每一跳都分析 IP 报头，节约了处理时间，提高了转发效率。

随着 ASIC（Application Specific Integrated Circuit）技术的发展，路由查找速度已经不是阻碍网络发展的瓶颈，这使得 MPLS 在提高转发速度方面不再具备明显的优势。但 MPLS 的设计者在一开始就充分吸取了 IP 和 ATM 的技术精华，将 MPLS 协议定位在网络体系结构中的第 2.5 层位置，即位于 TCP/IP 协议栈中的链路层和网络层之间，用于向 IP 层提供连接服务，同时又从链路层得到服务。其“multiprotocol”的设计理念，更使得 MPLS 协议可以承载多种网络层和链路层协议报文，所以 MPLS 在 VPN、TE（流量工程）和 QoS 等应用方面变得更加灵活。

1.1.2 MPLS 网络结构

MPLS 网络的典型结构如图 1-1 所示，网络中各路由器（也可以是三层交换机）称作 LSR（Label Switching Router，标签交换路由器）。由这些 LSR 构成的网络区域称为 MPLS 域（MPLS Domain），其中位于 MPLS 域边缘、连接其他网络（如 IP 网络）的 LSR 称为 LER（Label Edge Router，边缘路由器），MPLS 域内的 LSR 称为核心 LSR（Core LSR）。MPLS 域通常也称 MPLS/IP 骨干网，属于所有用户共享的底层公网，不同用户在这个 MPLS 公网上可以建立自己的私网，即 VPN 网络，这就是在配套的《华为 MPLS VPN 学习指南》一书中所介绍的各种 MPLS VPN 解决方案，可以是三层 VPN（L3VPN，如 BGP/MPLS IP VPN），也可以是二层 VPN（L2VPN，如 VLL、PWE3、VPLS 等）。

IP 报文在 MPLS 网络转发过程中所经过的路径称为 LSP（Label Switched Path，标签交换路径）。一条 LSP 可以看成是一条 MPLS 隧道，专用于一类 IP 报文的传输。其中的入口 LER 称为入节点（Ingress）；位于 LSP 中间的 LSR 称为中间节点（Transit）；LSP 的出口 LER 称为出节点（Egress），如图 1-1 所示。一条 LSP 可以有 0 个、1 个或多个中间节点，**但有且只有一个入节点和一个出节点**。根据 LSP 的方向，MPLS 报文由 Ingress 发往 Egress，则 Ingress 是 Transit 的上游节点，Transit 是 Ingress 的下游节点。同理，Transit 是 Egress 的上游节点，Egress 是 Transit 的下游节点。

【经验提示】因为 LSP 是单向的，要实现隧道两端所连网络的互通，仅一个方向的 LSP 是不行的，还需要建立相反方向、Ingress 节点和 Egress 节点角色互换的 LSP。一条

LSP 代表一条 MPLS 隧道，这也就是说 **MPLS 隧道也是单向的**。所以通常情况下，为了实现最终 MPLS 网络的互通，需要建立两条相反方向的 LSP，即两条相反方向的 MPLS 隧道。在 LER 设备上一般要同时进行 Ingress 和 Egress 两方面的配置，因为一台 LER 设备会同时担当 Ingress 和 Egress 角色，当然对应不同的 LSP。

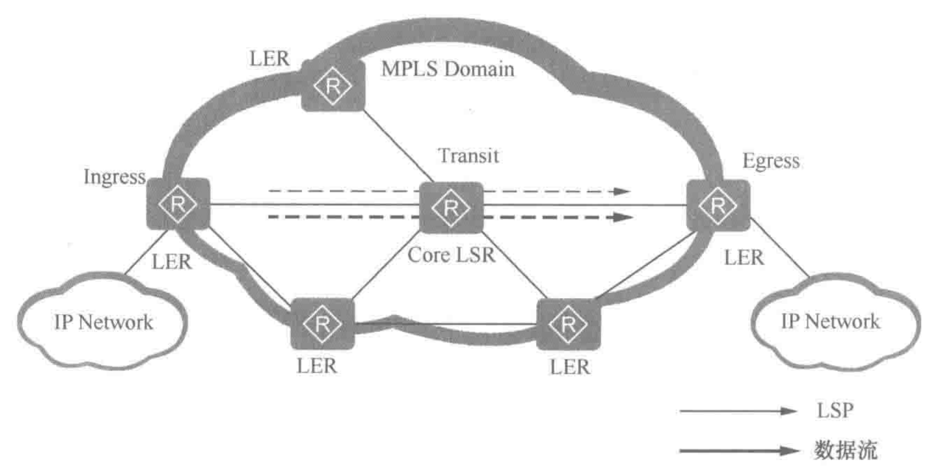


图 1-1 MPLS 网络典型结构

MPLS 的基本工作过程如图 1-2 所示。首先 MPLS/IP 骨干网中的设备会为隧道两端 LER 上连接的每个公网网段（称之为 FEC，转发等价类）建立一条 LSP，路径上的每台设备都会为该 LSP 分配一个用于指导 MPLS 报文转发的 MPLS 标签。该 MPLS 标签又与报文转发的下一跳和出接口相映射，使得 MPLS 报文在骨干网中传输时可以直接依据各设备上为该报文所分配的 MPLS 标签进行转发。但 MPLS 报文上的标签不是固定不变的，而是随着报文的传输，每经过一个设备都需要进行替换，以获得从当前设备向下游节点继续转发报文的路径。所以从本质上讲，MPLS 报文在骨干网中的转发过程实质上就是 MPLS 报文中 MPLS 标签的逐跳交换过程。

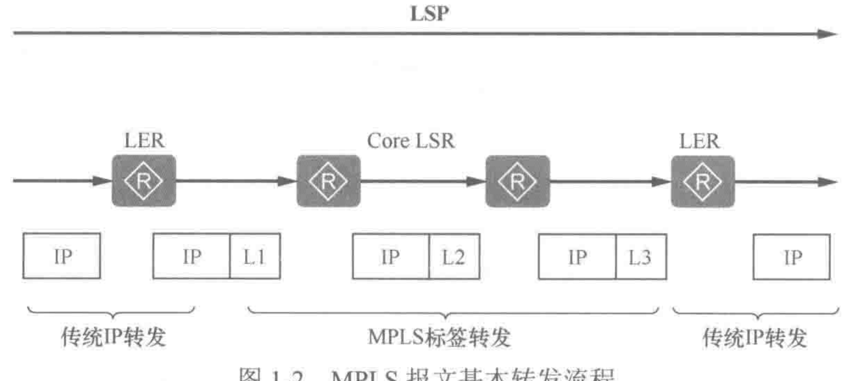


图 1-2 MPLS 报文基本转发流程

当 IP 报文进入 MPLS 域的 LER（此时为入节点）时，首先根据其目的 IP 地址找到的是其对应的 FIB（转发信息表）表项，如果其中的 Tunnel ID 值不为 0，则表示要进行 MPLS 标签转发（采用 IP 路由进行转发的表项对应的 Tunnel ID 值均为 0）。

在入节点进行 MPLS 转发前，需在 IP 报文的二层协议头和 IP 报头之间加上一层本

地设备为该LSP分配的MPLS标签（L1），然后根据标签L1映射的出接口及下一跳传输给下游的Core LSR。在Core LSR中的MPLS报文的标签也要先用本地设备为该LSP分配的标签（L2）替换MPLS报文中原来的标签（L1），然后再根据新标签L2所映射的出接口及下一跳进行转发（不用按照路由表进行转发）。继续按照同样的方法向下游节点转发，到了MPLS域另一端LER（出节点）时，通常会去掉MPLS报文中的MPLS标签，还原为普通IP报文，按照IP路由方式进行转发。

当然，以上介绍的仅是最基本的MPLS报文转发流程，其目的仅是想先让大家对MPLS标签交换的基本设计思想有一个初步了解，详细的转发流程将在1.2节介绍。

1.1.3　MPLS标签

MPLS标签（MPLS Label）是一个**短而定长**（这样开销可以很小），且只具有本地意义（无需全网唯一）的整数形式的数字标识符，用于唯一标识一个分组所属的分类（类似于IP路由中的Tag标记），这个分类称之为FEC（Forwarding Equivalence Class，转发等价类）。一个FEC中的分组数据在同一台设备上都将以等价（相同）的方式进行处理，且被分配相同的MPLS标签。

MPLS将具有相同特征的报文归为同一FEC，这个相同"特征"可以根据报文中的源IP地址、目的IP地址、源端口、目的端口、VPN实例、QoS策略等要素中的一个或多个进行划分，**但通常是根据目的IP地址基于IP路由表项划分**。例如，在传统的采用最长匹配原则的IP路由转发中，采用同一条路由的所有报文就是一个转发等价类FEC。在每台MPLS设备上，每个FEC与MPLS标签之间有一个映射关系，但针对同一FEC，不同设备上分配的标签可以相同，也可以不同。

1. MPLS标签封装

MPLS的应用比较广泛，在不同应用中MPLS标签嵌入的位置不完全相同。在大多数MPLS应用中（包括BGP/MPLS IP VPN、各种VLL和PWE3），MPLS设备从用户端设备接收数据帧后，会在原来数据帧中的二层协议头和三层协议头（通常为IP协议头）之间插入一个或多个MPLS标签（MPLS Lable），如图1-3所示。



图1-3　多数MPLS应用中MPLS标签在报文中封装的位置

在VPLS应用中，MPLS设备从用户端设备接收到数据帧后，会在原来数据帧中的二层协议头和新添加的二层协议头之间插入一层或多层MPLS标签，如图1-4所示。



图1-4　VPLS应用中MPLS标签在报文中封装的位置

说明　有关BGP/MPLS IP VPN、各种VLL、PWE3、VPLS方式的具体工作原理及配置与管理方法请参见《华为MPLS VPN学习指南》一书。

无论是哪种封装方式，一个 MPLS 标签（在一个 MPLS 报文中可能有多个 MPLS 标签）占 4 个字节（32 位），又包括多个子字段，如图 1-5 所示。具体含义说明如下。

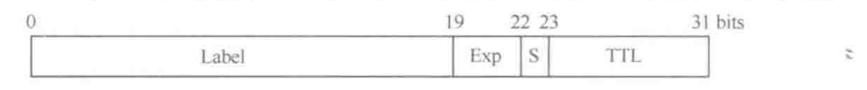


图 1-5 MPLS 标签结构

■ Label：20 bit，标签值字段，这是真正的 MPLS 标签取值部分，该字段的取值范围称之为“标签空间”，具体将在下面介绍。

■ Exp：3 bit，标识 MPLS 报文的优先级，即 MPLS 优先级，取值范围为 0～7 的整数。数值越小，优先级越低。当设备队列阻塞时，优先发送优先级高的报文。

■ S：1 bit，栈底标识。因为 MPLS 支持多层标签，即标签嵌套，为了识别哪个标签是 MPLS 报文中最后一个标签，用了这样一比特来进行标识。当栈底标签被弹出（剥离）时，则表示报文中不再携带 MPLS 标签，也就不再是 MPLS 报文了。S 值为 1 时表明该标签为最底层标签，其他各层标签该位为 0。

■ TTL：8 bit，和 IP 报文中的 TTL（Time To Live）意义相同，用于限制 MPLS 报文传输的距离，即最多能传输多少跳下游节点，当 TTL 值为 0 时，报文不能再下传输。该字段值初始化时有可能是 255，也有可能是从 IP 报头中的 TTL 字段复制得到的，具体参见本章 1.2.4 小节。

用于建立 LSP 隧道的 MPLS 标签有两种：入标签（In lable）和出标签（Out lable）。针对同一 LSP，MPLS 报文仅携带一层 MPLS 标签，可能是入标签（进入本地设备后），也可能是出标签（从本地设备发出时）。所谓“标签交换”是指 MPLS 报文从本地设备发出时，用本地设备为某 FEC 分配的出标签（也是下游节点为该 FEC 分配的入标签）替换报文中原来携带的本地设备为该 FEC 分配的入标签（即上游节点为该 FEC 分配的出标签），然后从出标签所映射的出接口转发出去。出标签是由下游节点为本地节点针对某 FEC 而分配的，与下游节点为该 FEC 分配的入标签一致。

每个属于同一 FEC 的 MPLS 报文进入一台设备时都会为该报文打上该设备为此 FEC 分配的入标签，然后找到与该入标签映射的出标签，继而找到对应的出接口，再把 MPLS 报文的入标签替换成所映射的出标签，从出接口发送出去。

MPLS 标签分发的方向与 LSP 方向是相反的，如图 1-6 所示。MPLS 标签最初是由目的 FEC 对应的 Egress 节点分配的，这时分配的也是某 FEC 在 Egress 设备上的入标签，然后再由该 Egress 设备向其上游节点进行通告（发送标签映射消息），上游节点收到这个通告后就把通告中的标签当作本地该 FEC 的出标签，依此类推。

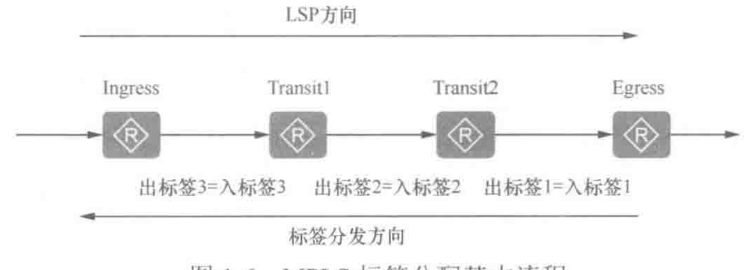


图 1-6 MPLS 标签分配基本流程

由此可知，MPLS 标签与 FR（帧中继）中的 DLCI 是类似的，也是要求相邻设备间

连接的接口所绑定的标签必须相同（即上游节点的出标签与本地设备的入标签必须相同）。所以，实际上，每跳设备仅需要为每个 FEC 分配一个标签，即入标签，但在 Ingress 节点可不分配入标签。

图 1-7 是一个 MPLS 网络中各节点所携带 MPLS 标签的示例。在静态 LSP 中，MPLS 入标签和出标签都是管理员手工配置的，而在由 LDP（Label Distribution Protocol，标签分发协议）等协议动态建立的 LSP 中，MPLS 标签是通过 LDP 协议自动分配的，具体将在本书第 3 章介绍。

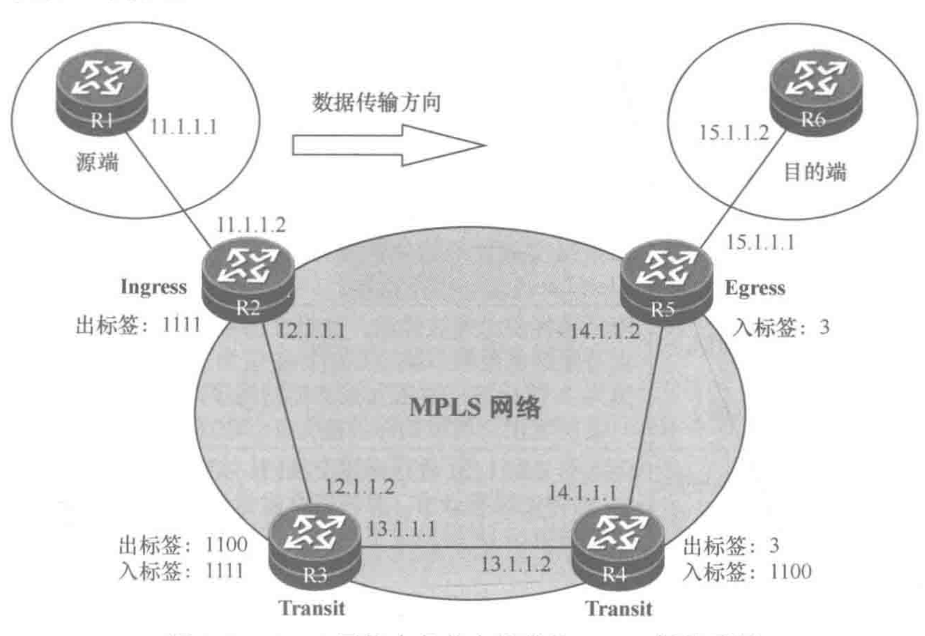


图 1-7　MPLS 网络中各节点携带的 MPLS 标签示例

2. MPLS 标签栈

如果 MPLS 报文中封装了多个 MPLS 标签（如既有 LDP LSP 标签，又有 BGP LSP 标签、MPLS CR-LSP 标签或 VC 标签等），就形成了标签栈（Label Stack）。图 1-8 所示的是一个包含有两个 MPLS 标签的示意图，靠近二层帧头的标签称为栈顶 MPLS 标签或外层 MPLS 标签（Outer MPLS label），此时 S 位（栈底标识）置 0；靠近三层报头的标签称为栈底 MPLS 标签或内层 MPLS 标签（Inner MPLS label），此时 S 位置 1。中间还可能有更多层次的 MPLS 标签。

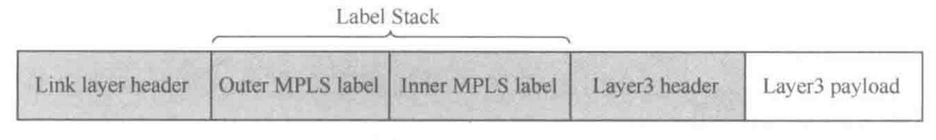


图 1-8　标签栈示意

指导数据转发的仅外层标签（如 LSP 标签），类似在 QinQ 中的多层 VLAN 标签中指导 VLAN 帧转发的仅外层 VLAN 标签一样，内层标签通常在 MPLS 网络传输途中不会发生变化，仅用于在到达出节点时查找报文转发的出接口，如 VC（虚电路）标签。理论上，MPLS 标签可以无限嵌套。

在图 1-5 所示的 MPLS 标签中的“Lable 字段”的取值范围（即标签空间），不同取

值的标签用途不一样，具体划分说明如下。

■ 0～15：这是 16 个特殊标签，具体说明见表 1-1。

表 1-1 特殊标签说明

标签值	含义	描述
0	IPv4 Explicit NULL Label（IPv4 显式空标签）	表示该标签必须被弹出（即标签被剥离），且报文的转发必须基于 IPv4。如果出节点分配给倒数第二跳节点的标签值为 0，则倒数第二跳 LSR 需要将值为 0 的标签正常压入报文标签栈顶部，转发给最后一跳。最后一跳发现报文携带的标签值为 0，则将标签弹出
1	Router Alert Label	只有出现在非栈底（即非最里层）的标签中才有效，因为栈底的标签会被直接弹出。类似于 IP 报文的“Router Alert Option”（路由器警报选项）字段，节点收到的 MPLS 报文中带有 Router Alert Label 标签时，需要将其送往本地软件模块（CPU 中对应的功能模块）进一步处理，实际报文转发由下一层（它的上层）标签决定。如果报文需要继续依据此标签进行转发，则节点需要将 Router Alert Label 压回标签栈顶
2	IPv6 Explicit NULL Label（IPv6 显式空标签）	表示该标签必须被弹出，且报文的转发必须基于 IPv6。如果出节点分配给倒数第二跳节点的标签值为 2，则倒数第二跳节点需要将值为 2 的标签正常压入报文标签栈顶部，转发给最后一跳。最后一跳发现报文携带的标签值为 2，则直接将标签弹出
3	Implicit NULL Label（隐式空标签）	倒数第二跳 LSR 进行标签交换时，如果发现交换后的出标签值为 3，则将该标签弹出，并将报文发给最后一跳。最后一跳收到该报文直接进行 IP 转发或下一层标签转发
4～13	保留	—
14	OAM Router Alert Label	MPLS OAM（Operation Administration & Maintenance，操作、管理和维护）通过发送 OAM 报文检测和通告 LSP 故障。OAM 报文使用 MPLS 承载。OAM 报文对于 Transit LSR 和倒数第二跳 LSR（penultimate LSR）是透明的，即不会对该标签进行交换和处理
15	保留	—

■ 16～1023：这是专门分配给静态 LSP 和应用于 MPLS TE 中的静态 CR-LSP（Constraint-based Routed Label Switched Path，基于约束的路由标签交换路径）共享的标签空间。

■ 1024 及以上：这是分配给 LDP、RSVP-TE（Resource Reservation Protocol-Traffic Engineering，资源保留协议流量工程）及 MP-BGP（MultiProtocol Border Gateway Protocol，多协议边界网关协议）等动态信令协议所分配的标签空间。即动态分配的 MPLS 标签号只能大于 1024。

1.1.4 MPLS 体系结构

MPLS 要实现标签的分配和交换，必须有一整套功能组件来完成，这就是 MPLS 的体系架构，如图 1-9 所示。

总体来说，MPLS 体系架构是由控制平面（Control Plane）和转发平面（Forwarding Plane）两部分组成。但在这两部分中，各自又包括了多个子项。

1．控制平面

控制平面用于控制协议报文的转发，其依靠 IP 路由和 MPLS 标签两方面来实现，因为

MPLS 骨干网中的 LSR 都是三层设备，需要依靠 IP 路由实现互通，而外部进入到 MPLS 骨干网的报文又要直接依靠 MPLS 进行转发，所以 LSR 的控制平面要同时负责对 IP 报文和 MPLS 报文转发的控制。要控制 IP 报文和 MPLS 报文转发就需要有产生、维护路由和标签信息的能力，这就是控制平面的基本功能，控制平面包括 3 个子项，各子项的职责如下。

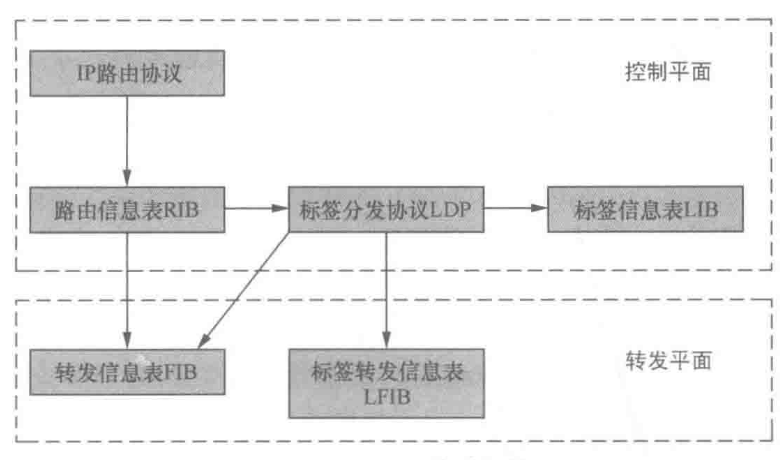


图 1-9　MPLS 体系架构

■ RIB（Routing Information Base，路由信息表）：RIB 由各种 IP 路由协议生成，用于进行路由选择。骨干网中 MPLS 标签的分发、LSP 的建立仍必须依据 IP 路由表，即先按照下面将要介绍的“转发信息表”（FIB）进行 MPLS 标签分发，建立对应的 LSP，然后 MPLS 域中的设备才可以直接按照标签进行数据转发。

■ LDP（标签分发协议）：LDP 是一种动态标签分发协议，负责 MPLS 标签的动态分发，LFIB（标签转发信息表）的建立，标签交换路径的建立、拆除等工作。**使用 LDP 协议进行标签分发的方向与 LSP 的方向相反**，具体将在本书第 3 章和第 4 章介绍。

■ LIB（Label Information Base，标签信息表）：MPLS 标签与 IP 路由中的 RIB 对应的就是 LIB，其由 LDP 生成，保存了每个标签与对应 FEC 的映射关系，用于管理 MPLS 标签信息。LIB 中包括的元素有：FEC 网段、入标签、出标签、分发出标签的下游节点等，它们之间建立了一一映射关系。

【经验提示】这个 LIB 包括了从本地设备到目的网段所有路径中获取的标签信息，但里面有些 LIB 表项在当前是无效的，即在当前不用于指导数据转发。类似动态路由协议中的拓扑数据库，拓扑数据库中存在到达某一目的地址的多条路径，但同一时刻一般只有一条是有效的（存在等价路由时可能同时存在多条有效路径）。保存所有标签信息的目的就是方便当网络拓扑结构发生变化时能及时调整 LSP。

每个 LSR 中都会基于所收到的每个 FEC 的标签映射信息，建立 LIB 表项。在这些相同或不同 LIB 表项中，不同标签之间的关系存在以下规则。

■ **所有的入标签必须不同**。因为入标签是由本地设备为不同 FEC 分配的，必须保证每个 FEC 所分配到的入标签唯一。但为同一 FEC 分配的入标签必须一致，不管其上游的路径有多少个。

■ **对于下一跳也相同的相同路由，出标签必须相同**。因为出标签是由下游节点分配的，当路由也相同（即同一 FEC）时，由同一下游节点所分配的出标签必须相同。

■ **对于下一跳相同的不同路由，出标签必须不同**。因为这是由同一个下游节点为不同 FEC 分配的出标签。

■ **对于下一跳也不同的不同路由，出标签可以相同，也可以不同**。因为这是不同下游节点为不同 FEC 所分配的路由，只需要在同一下游节点上保持为每个 FEC 所分配的入标签唯一即可。

■ **对于同一条路由，入标签和出标签可以相同，也可以不同**。同一设备上针对同一 FEC 上所映射的入标签是由本地设备分配的，出标签是由下游节点分配的，它们之间没有唯一性要求。

2. 转发平面

转发平面用于指导报文的转发，即数据平面（Data Plane）。其也包括 IP 报文转发和 MPLS 报文转发两个方面，负责构建各种用于指导 IP 报文、MPLS 报文转发的表项。指导报文的转发所需的就是出接口、下一跳这些基本元素。转发平面也包括两个子项，其职责如下。

■ FIB（Forwarding Information Base，转发信息表）：FIB 用于指导 IP 报文转告，是由从 RIB 提取必要的路由信息生成的，但仅提取当前有效的路由表项信息。当报文离开 MPLS 域时要按 FIB 进行转发。

FIB 中包括：目的网段、出接口、下一跳 IP 地址、路由标记、路由优先级等信息。在 FIB 中的表项都是当前有效的，如果过段时间，到达同一目的地址改变了所使用的路由表项，或者原来对应的路由表项被删除了，则原来的 FIB 表项也会自动删除，以确保里面的表项都可以在当时用于指导 IP 报文的转发。

■ LFIB（Label Forwarding Information Base，标签转发信息表）：LFIB 用于指导 MPLS 报文转发，由从 LIB 中提取必要的信息生成。LFIB 中除包括用于指导 IP 报文转发的目的网段、出接口、下一跳这三个基本元素外，还包括入标签和出标签。当 MPLS 报文在 MPLS 域内时需按 LFIB 进行转发。但 MPLS 中的 LFIB 与 IP 路由中的 FIB 类似，也仅包括 LIB 中当前有效的那些标签映射表项。

为了更直观地帮助大家理解以上各表中所包括的主要元素，现举一个简单的示例。如图 1-10 所示的是一个针对到达 FEC 10.0.0.0/8 目的网段，在控制平面和转发平面生成的各表项及相互关系的示例。

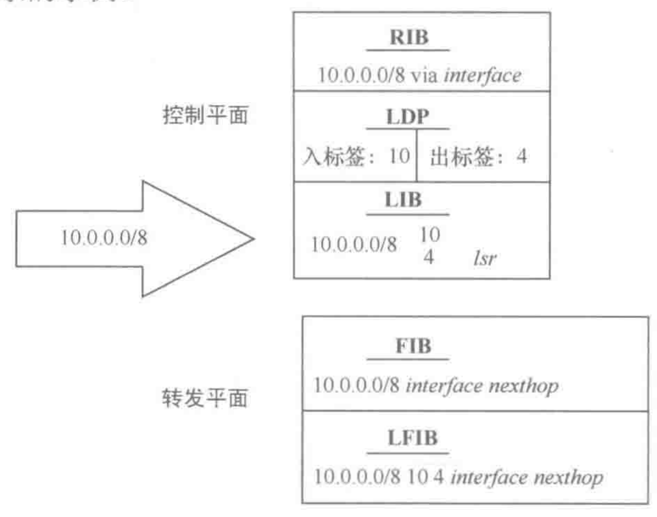


图 1-10 控制平面和转发平面各表项及相互关系示例

1.1.5 LSP简介

通过前面的学习已经知道，LSP是MPLS报文在MPLS网络中转发时经过的路径，可以看作是由报文传输方向各节点为对应FEC分配的MPLS入标签组成的，因为每台设备上为每个FEC分配的入标签是唯一的，并与由下游节点为本地节点上该FEC分配的出标签建立映射关系，所以入标签确定后即可确定唯一的转发路径。LSP仅用于指导报文从MPLS骨干网入节点（Ingress）到达出节点（Egress）之间转发的路径，不包括骨干网外的转发，所以LSP可以看成是入节点到出节点之间的通信路径。

LSP是由途经节点分段建立的，路径中各节点上为某FEC建立的LSP串联起来就是对应FEC的整条LSP。各节点上建立的LSP是由入标签，对应映射的出标签以及下一跳来确定转发路径的。像IP路由中从当前节点到达某目的网段可能有多条IP路由路径一样，在MPLS网络中从当前节点到达某FEC也可能会建立多条不同的转发路径（绑定多个不同的出标签、出接口和下一跳），但在同一时刻只有一条路径是最优、有效的。只有这条路径会进入到LFIB中。

MPLS中的LSP可以通过在各节点上静态配置标签来建立，也可以通过一些协议为节点动态分配标签来建立。静态LSP类似于静态路由，需要管理员在每个节点上分别手工配置，动态LSP相当于动态路由，是由标签分配协议为节点动态分配标签。下面分别予以介绍。

1. 静态LSP建立

静态LSP是管理员通过手工方式为各个转发等价类（FEC）分配标签而建立的，不需要标签分发协议参与，也不需要IP路由参与（但MPLS网络仍需部署路由，以实现骨干网内各LSR的三层互通）。由于静态LSP各节点上不能相互感知到整个LSP的建立情况，因此静态LSP是一个本地的概念，即本地LSP是否建立成功仅与本地设备对应端口的MPLS功能及状态有关。当然，最终还需要途经的各节点都建立好了基于某FEC的LSP，才能实现报文在MPLS网络中从入节点正确、成功地转发到出节点。

在静态LSP配置中，对于MPLS域中的不同节点所需配置的标签不一样。

- 对于入（Ingress）节点只需要配置出标签。
- 对于中间（Transit）节点需要同时配置入标签和出标签。
- 对于出（Egress）节点只需要配置入标签。

配置好静态LSP后，就相当于在设备上手动创建好了每个FEC的LIB和LFIB，而且一般情况下，LIB和LFIB中所包括的标签都是完全相同的，因为手工配置方式一般只配置真正用于报文转发的LSP，而不像动态LSP，通过标签分发协议会生成一些当前并无效的LSP。但要注意的是，LSP是单向的，所以如果需要两端能正常通信，源端和目的端的通信需要建立双向LSP，这两条LSP的Ingress和Egress角色是互换的。

静态LSP不使用标签发布协议，不需要交互控制报文，因此消耗资源比较小，适用于拓扑结构简单并且稳定的小型网络。但通过静态方式分配标签建立的LSP不能根据网络拓扑变化动态调整（就像静态路由一样），需要管理员干预。

有关静态LSP的配置方法将在第2章具体介绍。

2. 动态 LSP

动态 LSP 是通过标签发布协议（如 LDP、MP-BGP、RSVP-TE）动态建立的，但同时也需要 IP 路由参与，以便按照路由路径在相邻节点间彼此交换针对具体 FEC 的 MPLS 标签，实现由下游向上游分发 MPLS 标签，最终建立 LSP 的目的。不同的标签发布协议的 LSP 建立原理不一样，具体将在本书后面各章介绍。

3. 标签发布协议

MPLS 可以使用以下多种标签发布协议。

（1）LDP

LDP（Label Distribution Protocol，标签分发协议）是专为标签发布而制定的协议，是最常用的标签发布协议。LDP 根据 IGP（Interior Gateway Protocol，内部网关协议）及 BGP（Border Gateway Protocol，边界网关协议）对应的 IP 路由信息以逐跳方式建立 LSP。

有关 LDP 的具体工作原理和配置方法将在本书第 3～4 章介绍。

（2）RSVP-TE

RSVP-TE（Resource Reservation Protocol Traffic Engineering，资源预留协议流量工程）是对 RSVP（资源预留协议）的扩展，用于建立基于约束路由的 LSP（Constraint-based Routed Label Switched Paths，CR-LSP）。其拥有普通 LDP LSP 没有的功能，如发布带宽预留请求、带宽约束、链路颜色和显式路径等。

有关 RSVP-TE 的具体工作原理和配置方法将在本书第 5～7 章介绍。

（3）MP-BGP

MP-BGP（Multiprotocol Border Gateway Protocol，多协议边界网关协议）是在 BGP 协议基础上扩展的协议。MP-BGP 支持为 MPLS VPN 业务中私网路由和跨域 VPN 的标签路由分配 BGP LSP 标签。

有关 MP-BGP 的具体工作原理和配置方法将在《华为 MPLS VPN 学习指南》一书中介绍。

1.1.6 MPLS 的主要应用

MPLS 的应用主要体现在“基于 MPLS 的 VPN”和“基于 MPLS 的流量工程”这两个方面，下面分别简单介绍。

1. 基于 MPLS 的 VPN

传统 VPN 一般是通过 GRE、L2TP、PPTP、IPSec 等隧道协议来实现私有网络间数据在公网上的传送，而 MPLS LSP 是通过标签交换在运营商 MPLS/IP 骨干网中形成的隧道，数据报文不再经过封装或者加密，在安全性上类似于 FR（帧中继）网络的专用网，因此，用 MPLS 实现 VPN 具有天然的优势。

另外，MPLS 的 VPN 中的用户设备无需为 VPN 配置 GRE、L2TP 等隧道，网络时延被降到最低。基于 MPLS 的 VPN 通过 LSP 也可将运营商 IP 骨干网所连接的私有网络的不同分支联结起来，形成一个统一的网络，其还支持对不同 VPN 间的互通控制，实现精确的访问权限控制。

如图 1-11 所示是一个 MPLS VPN 的基本结构示意，各部分组成说明如下。

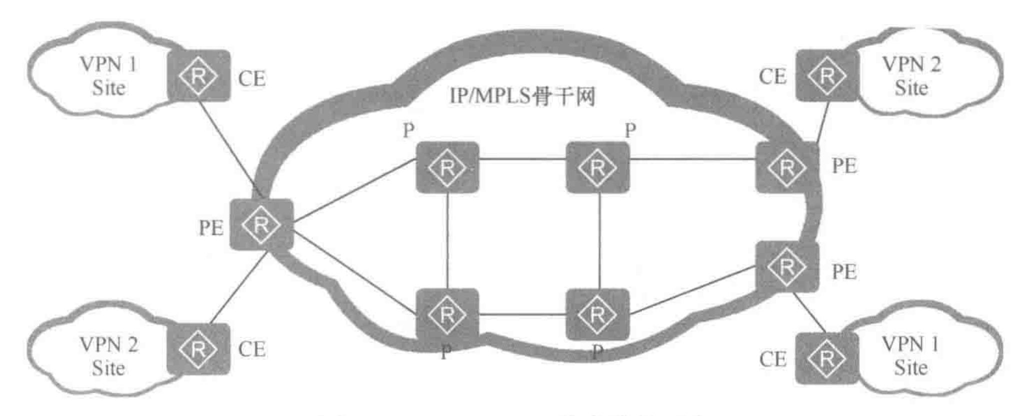


图 1-11 MPLS VPN 基本结构示意

■ CE（Customer Edge）：用户边缘设备，可以是路由器，也可以是交换机或主机。

■ PE（Provider Edge）：运营商边缘设备，是 MPLS/IP 骨干网的边缘设备。

■ P（Provider）： MPLS/IP 骨干网的核心设备，不与 CE 直接相连。P 设备只需要具备基本 MPLS 转发能力，不维护 VPN 信息。

基于 MPLS 的 VPN 有许多种，总体分为 L2VPN 和 L3VPN，L2VPN 主要包括 VLL（Virtual Leased Line，虚拟专线）、PWE3（Pseudo-Wire Emulation Edge to Edge，端到端伪线仿真）、VPLS（Virtual Private LAN Service，虚拟专用局域网业务）；L3VPN 主要有 BGP/MPLS IP VPN。这几种 VPN 都将在《华为 MPLS VPN 学习指南》一书中介绍。

基于 MPLS 的 VPN 具有以下特点。

■ PE 负责对 VPN 用户进行管理，建立各 PE 间 LSP 连接及同一 VPN 用户各分支间路由信息的发布。

■ PE 之间发布 VPN 用户路由信息通常通过 MP-BGP 协议实现。

■ 支持不同分支间 IP 地址复用和不同 VPN 间互通。

2. 基于 MPLS 的流量工程

传统的 IP 网络中，路由器选择最短的路径作为路由，不考虑带宽等因素。这样，即使某条路径发生拥塞，也不会将流量切换到其他的路径上。在网络流量比较小的情况下，这种问题不是很严重，但是随着互联网的发展及越来越广泛的应用，传统的最短路径优先的路由的问题暴露无遗。

TE（Traffic Engineering，流量工程）技术可通过动态监控网络的流量和网络单元的负载，实时调整流量管理参数、路由参数和资源约束参数等，使网络运行状态迁移到理想状态，从而优化网络资源的使用，避免负载不均衡导致的拥塞。

为了在大型骨干网络中部署流量工程，必须采用一种可扩展性好、简单的解决方案。MPLS 作为一种叠加模型，可以方便地在物理网络拓扑上建立一个虚拟拓扑，然后将流量映射到这个拓扑上。因此，基于 MPLS 的流量工程技术应运而生，即 MPLS TE。

如图 1-12 所示，从 LSR_1 到 LSR_7 存在两条路径：LSR_1→LSR_2→LSR_3→SR_6→SR_7 和 LSR_1→SR_2→SR_4→SR_5→SR_6→SR_7，现假设前者的带宽为 30Mbit/s，后者的带宽为 80Mbit/s。流量工程可以根据带宽等因素合理地分配流量，从而有效地避免链路拥塞。例如，LSR_1 到 LSR_7 存在两种业务，流量分别为 30Mbit/s 和 50Mbit/s，流量工程可以把前者分配到带宽为 30Mbit/s 的路径上，将后者分配到带宽为 80Mbit/s 的路

径上。

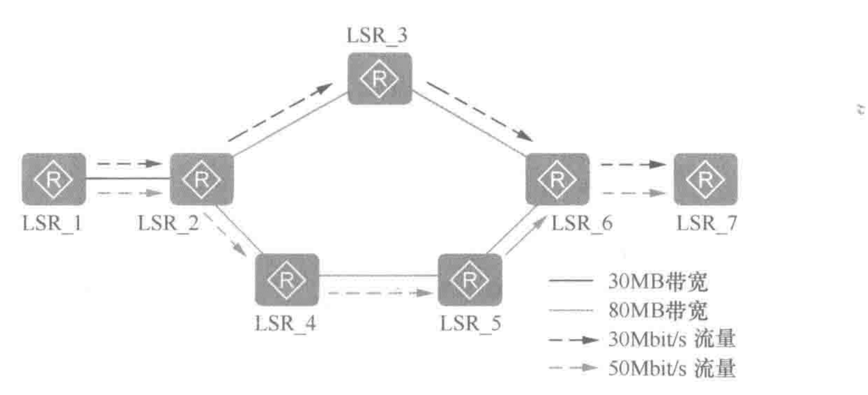


图 1-12 MPLS TE 示例

MPLS TE 通过建立经过指定路径的 LSP 进行资源预留，使网络流量绕开拥塞节点，达到平衡网络流量的目的。MPLS TE 具备以下优势。

- 在建立 LSP 隧道的过程中，可以为某种业务预留资源，以保证服务质量。
- LSP 隧道有优先级、带宽等多种属性，可以方便地控制 LSP 隧道的行为。
- 建立 LSP 隧道的负荷小，不会影响网络的正常业务。
- 通过备份路径和快速重路由技术，在链路或节点失败的情况下提供保护。

正是这些优势，使得 MPLS TE 成为流量工程的最佳方案。通过 MPLS TE 技术，服务提供商能够充分利用现有的网络资源，提供多样化的服务，同时可以优化网络资源，进行科学的网络管理。有关 MPLS TE 方面的技术原理、功能配置与管理方法以及应用将在本书第 5～7 章具体介绍。

1.2 MPLS 基本工作原理

MPLS 技术基本工作原理方面主要涉及 MPLS 标签动作、MPLS 报文转发流程，以及对 MPLS 报文中 TTL 的处理两个方面。

1.2.1 MPLS 标签动作

MPLS 基本转发过程中涉及一些标签操作，主要包括标签压入（Push）、标签交换（Swap）和标签弹出（Pop）这三个动作。

（1）Push：标签压入动作，可能会在 Ingress 或 Transit 节点上发生。

标签压入动作是指在 IP 报文的二层协议头和 IP 报头之间插入一个 MPLS 标签（如图 1-13 的上图所示），或者是在现有标签栈顶部再增加一个新的出标签（如图 1-13 的下图所示），即标签嵌套封装，如 BGP/MPLS IP VPN 的 Ingress 节点可能会在一个 IP 报文中同时压入多层公网或私网 MPLS 标签。

（2）Swap：标签交换动作，会在 Transit 节点发生。

当 MPLS 报文在 MPLS 域内转发时，Transit 节点根据标签转发表 LFIB 的查找，匹配到相应的表项后，用下一跳分配的出标签交换 MPLS 报文中原有的栈顶标签。原有

MPLS 报文中可以携带一层或多层 MPLS 标签，但仅交换最外层的标签。图 1-14 中上、下图所示分别是对携带单层标签和双层标签 MPLS 报文中的栈顶标签进行交换的示例。

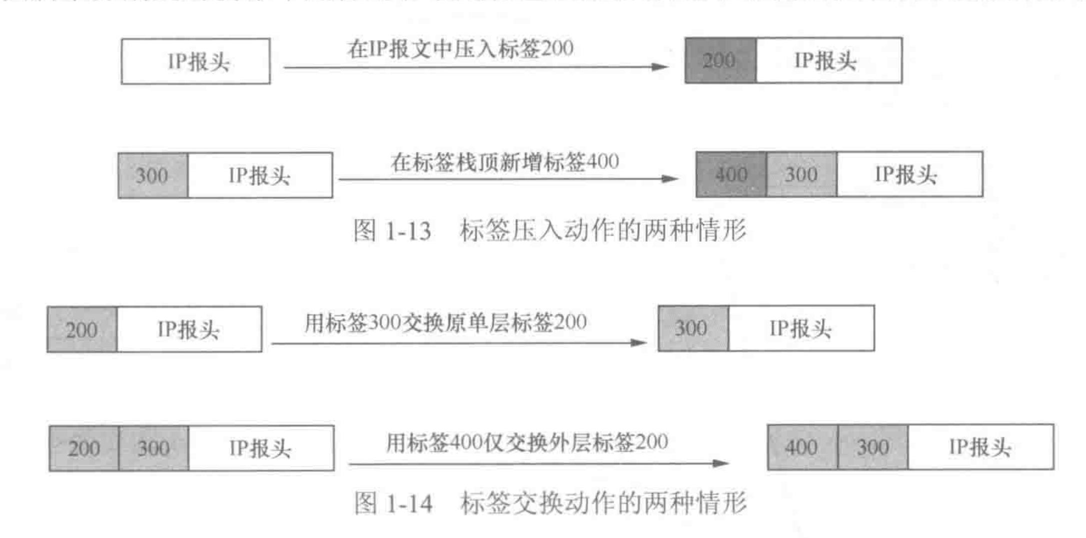


图 1-13　标签压入动作的两种情形

图 1-14　标签交换动作的两种情形

（3）Pop：标签弹出动作，会在倒数第二跳 Transit 节点或 Egress 节点发生。

当 MPLS 报文离开 MPLS 域时，Egress 节点将 MPLS 报文外层的标签剥离，使后续的报文转发按照 IP 路由进行（弹出标签后报文中无标签时，如图 1-15 中的上图所示），或者按照余下的标签进行（弹出标签后报文仍有其他标签时，如图 1-15 中的下图所示）。也可以利用 PHP（Penultimate Hop Popping，倒数第二跳弹出）特性，在倒数第二跳节点处将标签弹出，减少最后一跳的负担，使最后一跳节点直接进行 IP 路由转发或者下一层标签转发。

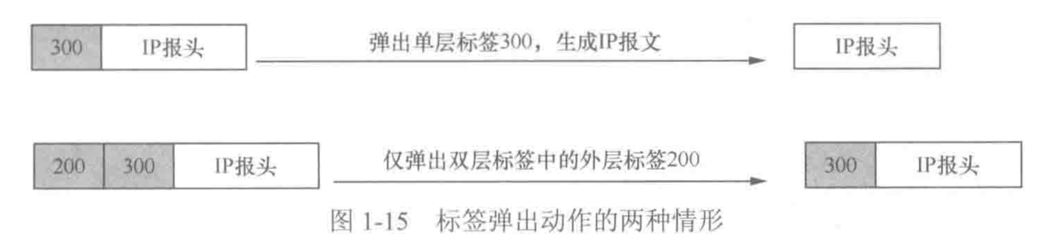


图 1-15　标签弹出动作的两种情形

默认情况下，华为设备支持 PHP 特性，支持 PHP 的 Egress 节点分配给倒数第二跳节点的标签值为 3。

以下以支持 PHP 的 LSP 为例，说明 MPLS 报文的基本转发过程。在单纯的 LDP LSP 隧道环境下，MPLS 报文最多仅带一层 MPLS 标签，从上游节点进入本地节点的入接口时携带的是上游节点分配给该 FEC 的出标签（也是本地节点对应的入标签），从本地节点出接口向下游节点发送时携带的是本地节点分配给对应 FEC 的出标签。

如图 1-16 所示，MPLS 标签已分发完成，建立了一条 LSP，其目的地址为 4.4.4.2/32，其 MPLS 报文的基本转发过程如下。

（1）Ingress 节点收到目的地址为 4.4.4.2 的 IP 报文后，首先根据 FIB 找到对应的下一跳，发现下一跳是 LSR 标签设备（如果发现下一跳是 IP 设备时会直接按 FIB 表项进行 IP 转发），并且因为本节点是入节点，所以在进行报文转发前需要进行标签压入动作，需压入的标签是根据 FEC 4.4.4.2 与标签的映射关系找到的（为 Z，作为出标签），然后

把 MPLS 报文从压入的标签所映射的出接口转发出去。

FEC	
4.4.4.2/32	
In/Out Label	In/Out IF
NULL/Z	IF1/IF2

FEC	
4.4.4.2/32	
In/Out Label	In/Out IF
Z/Y	IF1/IF2

FEC	
4.4.4.2/32	
In/Out Label	In/Out IF
Y/3	IF1/IF2

图 1-16　MPLS 报文基本转发示例

（2）Transit 节点收到该标签的 MPLS 报文后，根据 LFIB 找到对应入标签（上一节点的出标签就是本节点的入标签）所映射的出标签、出接口，先进行标签交换（无需查看 IP 报头的目的地址），即用本地为 FEC 4.4.4.2/32 分配的出标签（Y）替换报文中原来的 MPLS 标签（Z），然后从找到的出标签所映射的出接口转发出去。

（3）倒数第二跳 Transit 节点收到 MPLS 报文后，同样根据 LFIB 找到对应入标签所映射的出标签、出接口，先用本地为 FEC 4.4.4.2/32 分配的出标签（通常为 3）替换原来的 MPLS 标签，然后准备从出标签 3 所映射的出接口转发出去。但是因为 Egress 分给其的出标签值为 3（这是一个特殊的标签，必须弹出，参见表 1-1 说明），所以需要先进行 PHP 操作，弹出出标签（此时报文已不带 MPLS 标签了），并根据自己的出标签 3 所映射的接口转发报文。

（4）Egress 节点收到无 MPLS 标签的 IP 报文后，直接根据对应的 IP 路由表项把数据传输给目的主机 4.4.4.2/32。

1.2.2　MPLS 报文转发涉及的基本概念

1.2.1 节所介绍的只是 MPLS 报文的基本转发流程，具体的转发流程还涉及一些其他的技术细节，所涉及的相关概念如下。

（1）Tunnel ID

为了给使用隧道的上层应用（如 VPN、路由管理）提供统一的接口，系统会自动为隧道分配一个 ID（**在出节点上也可手动配置**），也称为 Tunnel ID。该 Tunnel ID 的长度为 32 比特，**只有本地意义**，即只要本地设备上唯一即可，同一条隧道中的不同节点的 Tunnel ID 可以一样。

（2）NHLFE

NHLFE（Next Hop Label Forwarding Entry，下一跳标签转发表项）用于指导 MPLS 报文的转发。NHLFE 包括：Tunnel ID、出接口、下一跳、出标签、标签操作类型等信息，可根据出标签找到对应的出接口及下一跳，进行报文转发。

FEC 与 NHLFE 的映射称为 FTN（FEC-to-NHLFE）。通过执行 **display fib** 命令查看

FIB 表中 Tunnel ID 值不为 0x0 的转发表项，能够获得 FTN 的详细信息。

FTN 只在 Ingress 存在，因为只在 Ingress 节点需要用到 FEC 中的分类信息来查找所需压入的出标签，然后再根据该出标签所映射的 NHLFE 找到对应的出接口及下一跳，进行报文的转发。后面的节点都是直接根据 MPLS 报文中所携带的出标签，在 NHLFE 中找到与出标签映射的出接口及下一跳信息进行报文转发。

（3）ILM

入标签与 NHLFE（下一跳标签转发表项）的映射称为 ILM（Incoming Label Map，入标签映射），其使本地设备的入标签和出标签、Tunnel ID 建立对应的关联关系。ILM 包括：Tunnel ID、入标签、入接口、标签操作类型等信息。

ILM 在 Transit 节点的作用是将入/出标签和 NHLFE 绑定。通过标签索引 ILM 表，就相当于使用目的 IP 地址查询 FIB，能够得到所有的标签转发信息。

下面以图 1-17 所示的示例介绍 MPLS 报文详细的转发过程。

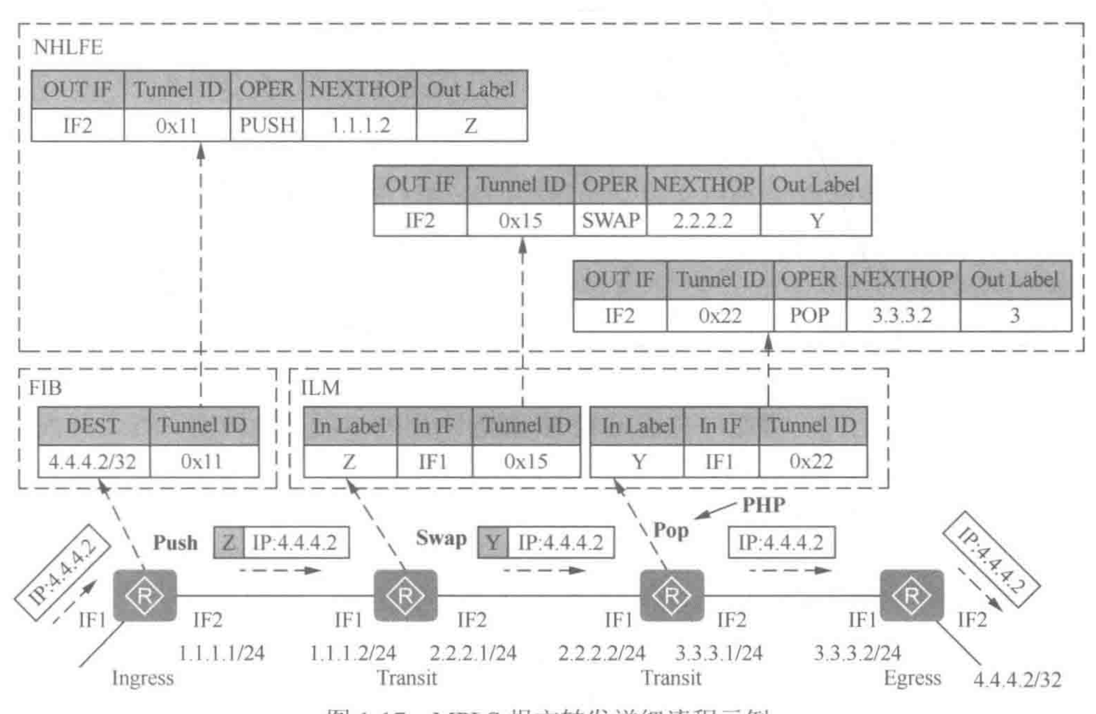


图 1-17　MPLS 报文转发详细流程示例

当 IP 报文从 Ingress 节点进入 MPLS 域时，首先查看 FIB 表，检查目的 IP 地址对应的 Tunnel ID 值是否为 0x0。

- 如果 Tunnel ID 值为 0x0，则进入正常的 IP 转发流程。
- 如果 Tunnel ID 值不为 0x0，则进入 MPLS 转发流程。

在 MPLS 转发过程中，FIB、ILM 和 NHLFE 表项都是通过 Tunnel ID 关联的。

1.2.3　MPLS 报文转发流程

MPLS 报文在骨干网中不同节点的具体转发流程有所不同，下面分别予以介绍。

1. Ingress 节点的转发流程

在 Ingress 节点上通过查询 FIB 表和 NHLFE 表指导报文的转发。

（1）首先根据 IP 报文的目的 IP 地址查看 FIB 表，即查看与目的 IP 地址对应的 Tunnel ID。如果需要采用 MPLS 转发，则 Tunnel ID 值肯定不为 0。

（2）然后根据 FIB 表的 Tunnel ID 找到对应的 NHLFE（下一跳标签转发表项），在 LSP 已建立的情况下，可在 NHLFE 查看到对应的出接口、下一跳、出标签和标签操作类型（此时为 Push）。

（3）在 IP 报文的二层协议头和三层 IP 报头之间压入一个出标签，同时处理 TTL，然后将封装好的 MPLS 报文发送给下一跳。

2. Transit 节点的转发流程

在 Transit 节点上通过查询 ILM（入标签映射）表和 NHLFE 表指导 MPLS 报文的转发。

（1）根据 MPLS 报文中的出标签值查看对应的 ILM 表（上游节点所压入的出标签与本地节点的入标签是相同的），可以得到对应的本地 Tunnel ID。

（2）然后再根据 ILM 表的 Tunnel ID 找到对应的 NHLFE 表项，可以得到进行下一跳转发所需的出接口、下一跳、出标签和标签操作类型。

（3）MPLS 报文的处理方式随着不同的标签值而不同。

■ 如果得到的出标签值≥16（表示该出标签不是特殊的标签），则用本地节点为该 FEC 分配的出标签替换原来 MPLS 报文中携带的标签（此时标签操作类型为 Swap），同时处理 TTL，然后将替换完标签的 MPLS 报文发送给下一跳。

■ 如果得到的出标签值为 3，则直接弹出 MPLS 报文中原来的出标签（此时标签操作类型为 Pop），同时处理 TTL，然后进行 IP 转发或下一层标签转发。

3. Egress 节点的转发流程

在 Egress 节点上，仅需通过查询 ILM 表来指导 MPLS 报文的转发，或通过查询 IP 路由表指导 IP 报文转发，因为出节点在 MPLS 域对应的 LSP 中没有下一跳设备，所以无需再利用 NHLFE 表来查询报文转发的出接口和下一跳。

（1）如果 Egress 收到的是不带 MPLS 标签的 IP 报文，则查看 IP 路由表，进行 IP 转发。

（2）如果 Egress 收到的是带有 MPLS 标签的 MPLS 报文，则查看 ILM 表获得的标签操作类型，同时处理 TTL。

■ 如果标签中的栈底标识 S=1，表明该标签是栈底标签，直接弹出该标签，然后进行 IP 转发。

■ 如果标签中的栈底标识 S=0，表明还有下一层标签（此时至少还有两层标签），继续进行下一层标签转发。

1.2.4　MPLS 对 TTL 的处理

MPLS 对 TTL 的处理包括 MPLS 对 TTL 的处理模式和 ICMP 响应报文这两个方面。

1. MPLS 对 TTL 的处理模式

MPLS 标签中包含一个 8 比特的 TTL 字段（参见图 1-5），其含义与 IP 报头中的 TTL 字段相同，用于限制报文在传输过程中所经过的三层设备数（每经过一跳值减 1）。MPLS 对 TTL 的处理与 IP 网络中对 TTL 的处理一样，除了用于防止产生路由环路外，也用于

实现 Traceroute（路由跟踪）功能。

RFC3443 中定义了两种 MPLS 对 TTL 的处理模式：Uniform（统一）和 Pipe（管道）。缺省情况下，MPLS 对 TTL 的处理模式为 Uniform。

（1）Uniform 模式

在 Uniform 模式下，针对 IP 报头中的 TTL 字段，最终在离开 MPLS 域时其值仍是减少了 MPLS 中所经过的跳数，只是在 MPLS 域内传输时，IP 报头 TTL 字段的改变移植到了 MPLS 标签中的 TTL 字段，然后在离开 MPLS 域时再反向移植到 IP 报头的 TTL 字段。

具体来说，IP 报文经过 MPLS 网络时，在入节点，IP 报头的 TTL 值减 1 映射到 MPLS 标签 TTL 字段，此后报文在 MPLS 网络中按照标准的 TTL 处理方式处理，即逐跳减 1，但 IP 报头中的 TTL 字段值在 MPLS 域中不改变。在出节点将 MPLS 标签的 TTL 减 1 后的值再映射到 IP 报头的 TTL 字段。最终的结果是，在 IP 报头的 TTL 字段值还是在逐跳减 1，即把 MPLS 网络中的每跳设备当作 IP 网络中的单跳来处理。

下面以图 1-18 所示的示例进行介绍。

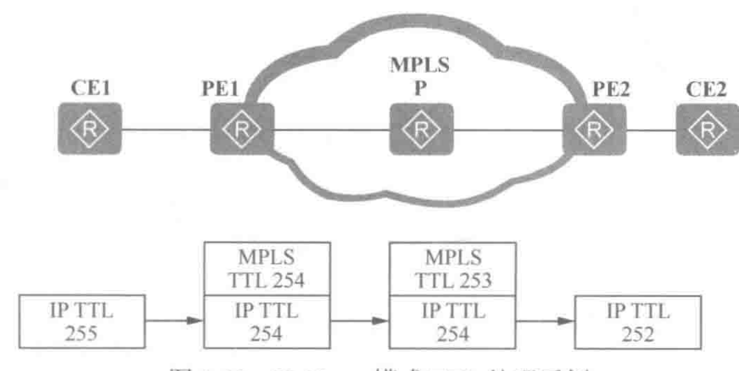


图 1-18　Uniform 模式 TTL 处理示例

1）IP 报文由 CE1 设备发出时，IP 报头中的 TTL 字段值为最大的 255。

2）到了 PE1 设备时，IP 报头中的 TTL 字段值减 1，变成了 254，然后把这个值映射到生成的 MPLS 标签 TTL 字段中，即 254。

3）MPLS 报文传输到中间设备 P 后，MPLS 标签 TTL 字段按标准 TTL 方式处理，即再减 1，变成了 253，但此时 IP 报头中的 TTL 字段值仍为 254，保持不变。

4）MPLS 报文继续传输到 PE2 设备时，MPLS 标签要弹出了，此时会先把 MPLS 标签中的 TTL 字段再减 1，得到 252，映射到 IP 报头中的 TTL 字段，使得到达 PE2 设备时 IP 报文的 TTL 字段值为 252。

从 CE1 到 PE2，总的来说经过了三跳设备，如果没有经过 MPLS 封装，IP 报头中的 TTL 字段值也要减 3，即 255-3=252，这与经过 MPLS 网络传输的结果是相同的，所以称之为 Uniform（统一）模式。

（2）Pipe 模式

在 Pipe 模式中，IP 报头中的 TTL 字段会把所经过的 MPLS 网络的中间节点忽略，将其看作两端边缘节点通过一个管道的直连（这也是称之为“Pipe”模式的原因），也就是无论在 MPLS 网络中经过了多少三层设备，从 MPLS 入节点到出节点之间，IP 报头中

的 TTL 字段值只减 1。

具体来说，IP 报文进入到 MPLS 网络的入节点后，IP 报头的 TTL 值减 1，MPLS 标签中的 TTL 字段为一个固定值（**这点与 Uniform 模式不同**），此后报文在 MPLS 网络中传输，MPLS 标签 TTL 字段值按照标准的 TTL 处理方式处理，即每经过一跳减 1，但 IP 报头的 TTL 值保持不变（**这点与 Uniform 模式相同**）。在出节点会直接弹出 MPLS 标签，然后将 IP 报头 TTL 字段的值减 1（**这点与 Uniform 模式也不同**）。即 IP 报文在经过 MPLS 网络时，无论中间经过了多少跳，IP 报头的 TTL 值只在入节点和出节点分别减 1，中间传输过程中保持不变，即相当于把 MPLS 网络中间节点忽略。

图 1-19 是一个采用 Pipe 模式的 TTL 处理示例。在 MPLS VPN 应用中，出于网络安全的考虑，需要隐藏 MPLS 骨干网络的结构，这种情况下，对于私网报文采用 Pipe 模式。因为这种模式下，MPLS 骨干网中间节点是看不出来的。

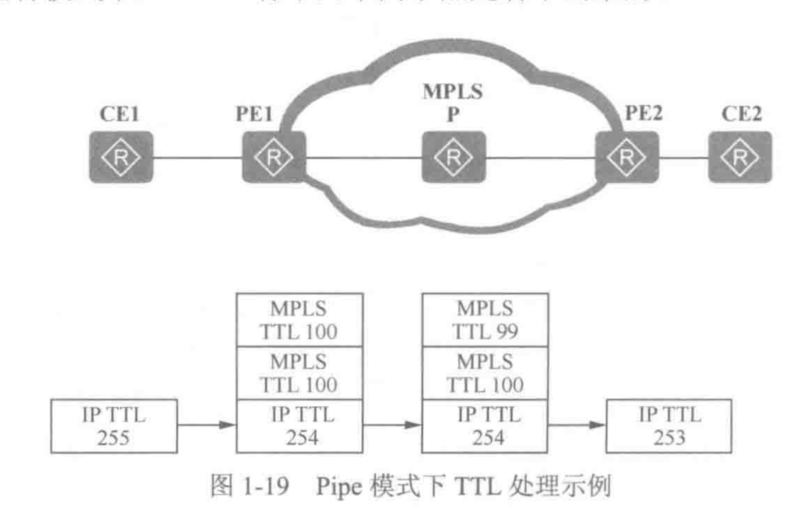


图 1-19 Pipe 模式下 TTL 处理示例

2. ICMP 响应报文

在 MPLS 网络中，当 LSR 收到 TTL 为 1（表示不能再向下传了）的含有标签的 MPLS 报文时，会生成一个 ICMP 的 TTL 超时消息。LSR 将 TTL 超时消息回应给报文发送者的方式有两种。

■ 如果 LSR 上存在到达报文发送者的路由，则可以通过 IP 路由，直接向发送者回应 TTL 超时消息。

■ 如果 LSR 上不存在到达报文发送者的路由，则 ICMP 响应报文将按照 LSP 继续传送，到达 LSP 出节点后，由 Egress 节点将该消息返回给发送者。

通常情况下，当收到的 MPLS 报文只带一层标签时，LSR 可以采用第一种方式回应 TTL 超时消息，因为此时表明 LSR 是 MPLS 域的边缘节点 LER，可直接通过 IP 路由传输响应报文。当收到的 MPLS 报文包含多层标签时，LSR 采用第二种方式回应 TTL 超时消息。但是，在 MPLS VPN 中，ASBR（Autonomous System Boundary Router，自治系统边界路由器）和 HoVPN（Hierarchy of VPN，分层 VPN）组网应用中的 SPE（Superstratum PE or Sevice Provider-end PE，上层 PE 或运营商侧 PE）接收到的承载 VPN 报文的 MPLS 报文也可能只有一层标签，此时，这些设备上并不存在到达报文发送者的路由，则采用第二种方法回应 TTL 超时消息。

1.3 LSP 连通性检测

在 MPLS 网络中，如果通过 LSP 转发数据失败，采用传统 IP 网络中的 Ping 或 Tracert 操作，负责建立 LSP 的 MPLS 控制平面将无法检测到这种错误，因为 ICMP 报文是基于 IP 路由转发的。此时，就要利用 RFC 4379 定义的 MPLS Ping/MPLS Tracert 功能来发现 LSP 错误，并及时定位失效节点。MPLS Ping 主要用于检查 LSP 的连通性，MPLS Tracert 在检查 LSP 的连通性的同时，还可以分析网络什么地方发生了故障，类似于普通 IP 网络中的 Ping/Tracert。

1.3.1 MPLS Ping/MPLS Tracert 简介

MPLS Ping/MPLS Tracert 使用 MPLS 回显请求（Echo Request）报文和 MPLS 回显应答（Echo Reply）报文检测 LSP 的可用性，这也与 IP 网络中 Ping/Tracert 所使用的回显请求和回显应答报文的工作机制类似。但很显然，MPLS 回显请求（Echo Request）报文和 MPLS 回显应答只有在使能了 MPLS 功能的设备才能被识别，因为这两种消息中都带有 MPLS 标签。

在 MPLS Ping 的 Echo Request/Echo Reply 报文中最外层 MPLS 出标签中的 TTL 字段为 255，而 MPLS Tracert Echo Request/Echo Reply 报文最外层 MPLS 出标签中的 TTL 字段值每次测试的赋值不同，依次是 1、2、3……这两种消息都以 UDP 报文格式发送，Echo Request 报文的目的 UDP 端口号为 3503，接收端通过 UDP 端口号识别出 MPLS Echo Request 报文，并发出 MPLS Echo Reply 报文进行和响应。

MPLS Ping/MPLS Tracert Echo Request 报文中携带需要检测的 FEC 信息（在请求报文中的“目的 FEC”字段中标识），和其他属于此 FEC 的报文一样沿 LSP 发送，从而实现对 LSP 的检测。Echo Request 报文以 MPLS 网络以标签为导向转发给目的端，而 Echo Reply 报文则以 IP 网络中的 IP 路由为导向转发给源端。另外为了防止 LSP 断路时，Echo Request 进行 IP 转发，保证 LSP 的连通性测试，将 Echo Request 消息 IP 头中目的地址设置为 127.0.0.1/8（本机环回地址，供 CPU 识别，被上送的报文需要由 CPU 自己处理），IP 报头中的 TTL 值为 1。

说明 在 MPLS Ping Echo Request 报文中，通常会将一个特殊的标签，比如 Router Alert Label（值为 1，参见表 1-1），置于用于转发的 MPLS 出标签后面（即 Router Alert Label 作为内层标签），当报文到达目的地之后，目的 LSR 看到 Router Alert Label 后会把该报文送到 CPU 处理。CPU 检查到报文中 IP 报头部分的目的 IP 地址是 127.0.0.1，就会对其进行处理。

下面以具体示例介绍这两种测试工具的基本工作原理。

1.3.2 MPLS Ping 工作原理

如图 1-20 所示，LSR_1 上建立了一条目的地为 LSR_4 环回接口所在网段的 LSP。

从 LSR_1 对 LSR_4 进行 MPLS ping 时的处理流程如下。

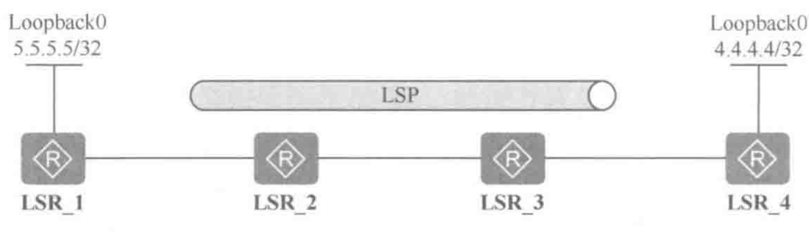


图 1-20 MPLS ping 测试示例

（1）执行 MPLS ping 命令后，LSR_1 首先会依据所 ping 的目的 IP 地址 4.4.4.4/32 在 FIB 中查看有没有对应的 TunnleID，由此判断该网段的 LSP 是否存在（对于 TE 隧道，查找 Tunnel 接口是否存在且 CR-LSP 是否建立成功）。如果不存在，返回错误信息，停止 ping 操作；如果存在，则继续进行以下步骤。

（2）在 LSR_1 上构造 MPLS Echo Request 报文，其中 IP 报头中的目的地址为 127.0.0.1/8，IP 报头中的 TTL 值为 1（**用于阻断 Echo Request 报文采用 IP 路由方式进行转发**），同时将 4.4.4.4 填入 Echo Request 报文中的“目的 FEC”字段中，然后根据对应的 NHLFE 查找相应的 LSP 的出标签，在 MPLS Echo Request 报文压入该出标签（内层中携带值为 1 的 Router Alter Lable 标签），将报文发送给 LSR_2。

（3）正常情况下，中间节点 LSR_2 和 LSR_3 对 MPLS Echo Request 报文会根据 MPLS Echo Request 报文中交换后的出标签进行普通的 MPLS 转发。如果中间节点 MPLS 转发失败，则中间节点返回带有错误码的 MPLS Echo Reply 报文，而这个回显应答报文不再通过 MPLS 标签转发，而是通过 IP 路由进行转发。

（4）当 MPLS 转发路径无故障时，则会把 MPLS Echo Request 报文送达 LSP 的出节点 LSR_4，然后检查目的 FEC 中包含的目的地址 4.4.4.4 是否为自己的 Loopback 接口地址，以此确定 LSR_4 是该 FEC 的真正出口后，上送到 CPU。CPU 看到该报文的 IP 报头中的目的 IP 地址为 127.0.0.1，则认为此请求报文需要由 CPU 自己处理，于是产生一个 MPLS Echo Reply 报文进行响应（同样采取 IP 路由方式转发）。至此整个 MPLS ping 过程结束。

1.3.3 MPLS Tracert 工作原理

同样以图 1-20 为例进行，从 LSR_1 对 4.4.4.4/32 进行 MPLS Tracert 时的处理如下。

（1）执行 MPLS Tracert 命令与执行 MPLS ping 命令一样，LSR_1 首先也会检查目的网段 4.4.4.4/32 的对应 LSP 是否存在（对于 TE 隧道，查找 Tunnel 接口是否存在且 CR-LSP 是否建立成功）。如果不存在，返回错误信息，停止 Tracert，否则继续进行如下处理。

（2）LSR_1 构造 MPLS Echo Request 报文，IP 报头中的目的地址为 127.0.0.1/8，同时将 4.4.4.4 填入 MPLS Echo Request 报文中的目的 FEC 中，然后从 NHLFE 中查找对应的 LSP，压入相应的 LSP 出标签，并且将 MPLS TTL 字段值设置为 1，将报文发送给 LSR_2。此 MPLS Echo Request 报文中包含 Downstream Mapping TLV（用来携带 LSP 在当前节点的下游信息，主要包括下一跳地址、出标签等）。

【经验提示】这里首先将 MPLS Echo Request 报文中的 MPLS TTL 字段值设置为 1

的目的其实与在 IP 网络中执行 Tracert 命令时进行第一跳测试时将 TTL 值设为 1 一样，就是用来进行第一跳的测试。第一跳测试成功后再进行第二跳、第三跳……的测试，直到到达目的端对应的 MPLS Echo Request 报文中的 MPLS TTL 字段值分别为 2、3……

（3）LSR_2 收到 LSR_1 发送来的 MPLS Echo Request 报文后，将报文中的 MPLS TTL 减 1 为 0 后发现 TTL 超时，然后 LSR_2 需要检查是否存在该 LSP，同时检查报文中 Downstream Mapping TLV 的下一跳 IP 地址、出标签是否正确。如果两项检查都正确，返回正确的 MPLS Echo Reply 报文（以 IP 路由方式转发），并且报文中必须携带 LSR_2 本身包含的下一跳和出标签的 Downstream Mapping TLV 给 LSR_1（**这个很重要，也是必须的，这是后面进行继续下一跳测试的依据**）。如果检查不正确，则返回错误的 MPLS Echo Reply 报文（也以 IP 路由方式转发）。

（4）LSR_1 收到正确的 MPLS Echo Reply 报文后再次发送 MPLS Echo Request 报文，报文的封装方式跟步骤（2）类似，只是将标签的 MPLS TTL 设置为 2，此时，MPLS Echo Request 报文中的 Downstream Mapping TLV 是从 MPLS Echo Reply 报文中复制过来的。然后 LSR_2 收到该报文后按出标签普通 MPLS 转发。LSR_3 收到此报文，标签的 TTL 超时，用与步骤（3）同样的处理后返回 MPLS Echo Reply 报文。

（5）LSR_1 收到正确的 MPLS Echo Reply 报文后重复步骤（4），把标签的 MPLS TTL 设置为 3，复制 Downstream Mapping TLV 后发送 MPLS Echo Request 报文。LSR_2 和 LSR_3 对该报文进行普通 MPLS 转发。LSR_4 收到此报文，重复步骤（3）处理方式对报文进行处理，同时检查目的 FEC 中包含的目的 IP 4.4.4.4 为自己的 Loopback 接口地址，以此来发现已经是该 LSP 的出节点，于是上送到 CPU，在上送的请求报文中发现 IP 报头中的目的 IP 地址为 127.0.0.1，因为需要要由 CPU 自己处理，生成一个不带下游信息的 MPLS Echo Reply 报文进行响应，至此整个 MPLS Tracert 过程结束。

通过上述步骤返回携带下游信息的 MPLS Echo Reply 报文，在 LSR_1 上就获取了该 LSP 沿途每一个节点的信息。

第2章 静态LSP配置与管理



静态 LSP 是 MPLS LSP 中最简单的一种 LSP，通过手工指定 MPLS 标签（无需信令协议分配 MPLS 标签）、目的 IP 地址、下一跳 IP 地址等参数，静态配置一条固定的 MPLS 隧道路径。与 IP 路由中的静态路由一样，LSP 路径参数都是手工静态指定的，故配置工作量比较大，容易出错，仅适用于小型 MPLS 骨干网中的 LSP 建立。

第 2 章专门介绍静态 LSP、基于静态 LSP 的 BFD 检测的配置与管理方法。最后介绍了静态 LSP 不通这种典型故障的基本排除方法，并在其中强调了静态 LSP 配置过程中一些特别要注意的问题。

2.1 静态 LSP 配置与管理

一般情况下，MPLS 网络中都使用 LDP 建立 LSP。但 LDP 是通过 IP 路由信息来建立 LSP 的，如果 LDP 协议出现问题，可能导致 MPLS 流量的丢失。因此，对于某些关键数据或重要业务，通过配置静态 LSP 来确定传输路径更为可靠。

静态 LSP 的优点是不使用标签发布协议，不需要交互控制报文，资源消耗比较小；缺点是通过静态方式建立的 LSP 不能根据网络拓扑变化动态调整，且需要管理员一条条手动配置，所以适用于拓扑结构简单、规模比较小、并且稳定的网络。

配置静态 LSP 时要遵循以下原则：**根据数据传输方向，上游节点 MPLS 出标签的值等于下游节点 MPLS 入标签的值**。但在不同类型节点上的配置不完全一样。

- 入节点需要指定 LSP 的目的 IP 地址（通常是 LSP 出节点担当 LSR-ID 的 Loopback 接口 IP 地址）和下一跳（可选同时配置出接口），但只需配置出标签。
- 中间节点需要配置入接口和下一跳（可选同时配置出接口），以及入标签和出标签。
- 出节点需要配置入接口和入标签。

要实现源和目的端相互通信，需要分别以两端 LER 为出节点创建双向静态 LSP。

2.1.1 创建静态 LSP

静态 LSP 的创建包括以下主要配置任务：配置 LSR ID→使能 MPLS→建立静态 LSP，使用的标签空间为 16～1023，具体配置步骤见表 2-1。但在创建静态 LSP 之前，也需要配置单播静态路由或 IGP，保证各 LSR 在网络层互通，以便在创建静态 LSP 时所指定的下一跳是可达的。

表 2-1 配置静态 LSP 的步骤

步骤	命令	说明
1	**system-view** 例如：<Huawei> **system-view**	进入系统视图
		配置 MPLS LSR ID
2	**mpls lsr-id** *lsr-id* 例如：[Huawei] **mpls lsr-id** 1.1.1.1	配置本节点的 LSR ID，用于唯一标识一个 LSR，点分十进制格式（与 IPv4 地址格式一样，类似于 OSPF、BGP 路由器 ID）。 在网络中部署 MPLS 业务时，必须首先配置 LSR ID，因为 **LSR 没有缺省的 LSR ID，必须手工配置**。为了提高网络的可靠性，推荐（**只是推荐，可以直接配置为其他 IPv4 地址**

（续表）

步骤	命令	说明
2	**mpls lsr-id** *lsr-id* 例如：[Huawei] **mpls lsr-id** 1.1.1.1	**格式的 LSR ID**）使用 LSR 某个 Loopback 接口的地址作为 LSR ID。建议 LSR ID 与 OSPF 或 BGP 的 Router ID 配置一样，整个网络唯一，用于对设备进行区分。 缺省情况下，没有配置 LSR ID，可用 **undo mpls lsr-id** 命令删除 LSR 的 ID。但如果要修改已经配置的 LSR ID，必须先在系统视图下执行 **undo mpls** 命令，然后再使用本命令配置
	使能 MPLS	
3	**mpls** 例如：例如：[Huawei] **mpls**	全局使能本节点的 MPLS，并进入 MPLS 视图。 缺省情况下，节点的 MPLS 能力处于未使能状态，可用 **undo mpls** 命令去使能全局 MPLS 功能，删除所有 MPLS 配置（除 LSR ID 外）
4	**quit** 例如：[Huawei] **quit**	返回系统视图
5	**interface** *interface-type interface-number* 例如：[Huawei] **interface** gigabitethernet 1/0/0	进入需要转发 MPLS 报文的接口的视图，**必须是三层接口，且必须是 MPLS 节点间相连的接口**
6	**mpls** 例如：[Huawei-GigabitEthernet1/0/0] **mpls**	使能以上接口的 MPLS。在需要部署 MPLS 业务的网络中，在节点上全局使能 MPLS 后，还需要在接口上使能 MPLS，才能够进行 MPLS 的其他配置。 缺省情况下，接口的 MPLS 能力处于未使能状态，可用 **undo mpls** 命令去使能接口的 MPLS 功能，删除所在接口的 MPLS 配置（包括接口下所有的 MPLS 配置）
	建立静态 LSP	
7	**static-lsp ingress** *lsp-name* **destination** *ip-address* { *mask-length* | *mask* } { **nexthop** *next-hop-address* | **outgoing-interface** *interface-type interface-number* } * **out-label** *out-label* 例如：[Huawei] **static-lsp ingress** staticlsp1 **destination** 10.1.0.0 16 **nexthop** 10.1.1.2 **out-label** 100	（三选一）在 Ingress 节点上配置静态 LSP。主要配置目的 IP 地址、下一跳 IP 地址（可同时配置出接口）和出标签。命令中的参数说明如下。 ● *lsp-name*：指定 LSP 名称（注意，不是 LSR ID），字符串形式，**区分大小写**，不支持空格，长度范围是 1～19。当输入的字符串两端使用双引号时，可在字符串中输入空格。 ● **destination** *ip-address*：指定目的 IP 地址。 ● *mask-length* | *mask*：指定目的 IP 地址所对应的子网掩码长度或子网掩码。 ● **nexthop** *next-hop-address*：可多选参数，指定下一跳 IP 地址，如果是以太网链路，则必须配置下一跳 IP 地址。 ● **outgoing-interface** *interface-type interface-number*：可多选参数，指定 LSP 的出接口。 ● **out-label** *out-label*：指定出标签值，整数形式，取值范围是 16～1023。 推荐采用指定下一跳的方式配置静态 LSP，确保本地路由表中存在与指定目的 IP 地址精确匹配的路由项，包括目的 IP 地址和下一跳 IP 地址 缺省情况下，没有为入节点配置静态 LSP，**可用 undo static-lsp ingress** *lsp-name* 命令为入节点删除一条 LSP，**但需要修改配置时，可直接重新配置，而不用先删除原来的配置**

（续表）

步骤	命令	说明
7	**static-lsp transit** *lsp-name* **incoming-interface** *interface-type interface-number* **in-label** *in-label* { **nexthop** *next-hop-address* | **outgoing-interface** *interface-type interface-number* } * **out-label** *out-label* 例如：[Huawei] **static-lsp transit** bj-sh **incoming-interface** gigabitethernet 1/0/0 **in-label** 123 **nexthop** 202.34.114.7 **out-label** 253	（三选一）在 Transit 节点上配置静态 LSP。主要配置入接口、入标签（**与上游节点配置的出标签要一致**），下一跳 IP 地址（可同时配置出接口）和出标签。 命令中的参数与在 Ingress 节点上配置 LSP 的参数类似，只是这里要同时配置入接口/入标签（**incoming-interface**/**in-label**）、下一跳、出接口/出标签（**nexthop**、**outgoing-interface**/**out-label**），入标签与出标签的取值范围也一样，为 16~1023。 推荐采用指定下一跳的方式配置静态 LSP，确保本地路由表中存在与指定目的 IP 地址精确匹配的路由项，包括目的 IP 地址和下一跳 IP 地址。如果 LSP 出接口为以太网类型，必须配置 **nexthop** *next-hop-address* 参数以保证 LSP 的正常转发。 缺省情况下，没有为中间转发节点配置静态 LSP，可用 **undo static-lsp transit** *lsp-name* 命令为中间转发节点删除一条 LSP，**但需要修改配置时，可直接重新配置，而不用先删除原来的配置**
	static-lsp egress *lsp-name* **incoming-interface** *interface-type interface-number* **in-label** *in-label* [**lsrid** *ingress-lsr-id* **tunnel-id** *tunnel-id*] 例如：[Huawei] **static-lsp egress** bj-sh **incoming-interface** gigabitethernet 1/0/0 **in-label** 233	（三选一）在 Egress 节点上配置静态 LSP。主要配置入接口、入标签（**与倒数第二跳节点配置的出标签要一致**）。可选参数 **lsrid** *ingress-lsr-id* **tunnel-id** *tunnel-id* 分别用来指定本地 LSR 的 LSR ID（在本表第 2 步配置的）和隧道 ID（取值范围是 1～65535）。 缺省情况下，没有在出节点配置静态 LSP，可用 **undo static-lsp egress** *lsp-name* 命令在出节点删除配置的静态 LSP。如果要修改 **incoming-interface** *interface-type interface-number*、**in-label** *in-label* 参数，可先删除原本的 LSP，只需重新执行本命令配置即可

【经验提示】从上表的静态 LSP 配置可以看出，只有 Ingress 才需要配置目的 IP 地址（相当于进行 FEC 划分），在 Transit 和 Egress 上均无需配置目的 IP 地址。所以为了确保各设备配置的静态 LSP 能完整体现对应 FEC 的整条 LSP，建议各设备上针对同一 FEC 配置的静态 LSP 名称相同。

另外，对于同一设备的 LSP，入标签和出标签可以是相同的，但上游节点的出标签值必须与下游节点的入标签相同。**对于同一设备上不同 LSP，在同一设备上所分配的入标签必须不同。**

2.1.2 配置静态 BFD 检测静态 LSP

这是一项可选配置任务，通过配置静态 BFD 检测静态 LSP，可以检测静态 LSP 的连通性，**需要在入节点和出节点同时配置**。配置静态 BFD 检测静态 LSP 时，需注意以下事项（有关 BFD 方面的技术原理参见《华为路由器学习指南》）。

■ 对非主机路由也可以建立 BFD 会话。当静态 LSP 的状态变为 Down 时，BFD 会话的状态也变为 Down；当静态 LSP 的状态变为 Up 时，会重新建立 BFD 会话。

■ **往返转发方式可以不一致**（如报文从源端到目的端使用 LSP 转发，从目的端到源端使用 IP 转发），**但往返路径要一致**。如果不一致，则检测到故障时，不能确定具体

是哪条路径的故障。

1. 配置入节点 BFD 参数

入节点可配置的 BFD 参数包括：所绑定的本地静态 LSP、本地标识符、远端标识符、本地发送 BFD 报文的时间间隔、本地接收 BFD 报文的时间间隔和本地 BFD 检测倍数，这些将会影响会话的建立。用户可以根据网络的实际状况调整本地检测时间。对于不太稳定的链路，如果本地检测时间较小，则 BFD 会话可能会发生震荡，这时可以选择延长本地检测时间。入节点的 BFD 参数配置步骤见表 2-2。

表 2-2 入节点 **BFD** 参数的配置步骤

步骤	命令	说明
1	**system-view** 例如：<Huawei> **system-view**	进入系统视图
2	**bfd** 例如：[Huawei] **bfd**	对本节点使能全局 BFD 能力并进入 BFD 全局视图。 缺省情况下，全局 BFD 功能未使能，可用 **undo bfd** 命令全局去使能 BFD 功能，此时如果已经配置了 BFD 会话信息，则所有的 BFD 会话都会被删除
3	**quit** 例如：[Huawei-bfd] **quit**	返回系统视图
4	**bfd** *cfg-name* **bind static-lsp** *lsp-name* 例如：[Huawei] **bfd bfd** 1to4 **bind static-lsp** 1to4	配置 BFD 会话所绑定的静态 LSP。命令中的参数说明如下。 • *cfg-name*：指定 BFD 配置名，字符串形式，不支持空格，不区分大小写，长度范围是 1～15。当输入的字符串两端使用双引号时，可在字符串中输入空格。 • *lsp-name*：指定 BFD 会话绑定静态 LSP 的名称，必须是已在上节表 2-1 第 7 步中所创建的静态 LSP 名称。 缺省情况下，没有创建检测静态 LSP 的 BFD 会话，可用 **undo bfd cfg-name** 命令删除指定的 BFD 会话
5	**discriminator local** *discr-value* 例如：[Huawei-bfd-session-1to4] **discriminator local** 10	配置本地标识符，整数形式，取值范围是 1～8191。 BFD 会话两端设备的本地标识符和远端标识符需要分别对应，即本端的本地标识符与对端的远端标识符相同，否则会话无法正确建立。并且，本地标识符和远端标识符配置成功后不可修改，如果需要修改静态 BFD 会话本地标识符或者远端标识符，则必须先删除该 BFD 会话，然后再配置本地标识符
6	**discriminator remote** *discr-value* 例如：[Huawei-bfd-session-1to4] **discriminator remote** 20	配置远端标识符，整数形式，取值范围是 1～8191
7	**min-tx-interval** *interval* 例如：[Huawei-bfd-session-1to4] **min-tx-interval** 300	（可选）调整本地发送 BFD 报文的时间间隔，整数形式，取值范围是 10～2000，单位是毫秒。 如果 BFD 会话在设置的检测周期内没有收到对端发来的 BFD 报文，则认为链路发生了故障，BFD 会话的状态将会置为 Down。为降低对系统资源的占用，一旦检测到 BFD 会话状态变为 Down，系统自动将本端的发送间隔调整为大于 1000 毫秒的一个随机值，当 BFD 会话的状态重新变为 Up 后，再恢复成用户配置的时间间隔。

（续表）

步骤	命令	说明
7	**min-tx-interval** *interval* 例如：[Huawei-bfd-session-1to4] **min-tx-interval** 300	【说明】用户可以根据网络的实际状况延长或者缩短 BFD 报文的发送和接收时间间隔。BFD 报文的发送、接收时间间隔直接决定了 BFD 会话的检测时间。对于不太稳定的链路，如果配置的 BFD 报文的发送、接收时间间隔较小，则 BFD 会话可能会发生震荡，这时可以选择延长 BFD 报文的发送和接收时间间隔。通常情况下，建议使用缺省值。 缺省情况下，发送间隔是 1000 毫秒，可用 **undo min-tx-interval** 命令恢复 BFD 报文的发送间隔为缺省值
8	**min-rx-interval** *interval* 例如：[Huawei-bfd-session-1to4] **min-rx-interval** 600	（可选）调整本地接收 BFD 报文的时间间隔，整数形式，取值范围是 10～2000，单位是毫秒。 缺省情况下，接收间隔是 1000 毫秒，可用 **undo min-rx-interval** 命令恢复 BFD 报文的接收间隔为缺省值
9	**detect-multiplier** *multiplier* 例如：[Huawei-bfd-session-1to4] **detect-multiplier** 5	（可选）调整本地 BFD 检测倍数，整数形式，取值范围是 3～50。 BFD 会话的本端检测倍数直接决定了对端 BFD 会话的检测时间，检测时间=接收到的远端 Detect Multi×max（本地的 RMRI，接收到的 DMTI），其中，Detect Multi 是检测倍数，通过本条命令配置；RMRI 是本端能够支持的最短 BFD 报文接收间隔；DMTI 是本端想要采用的最短 BFD 报文的发送间隔。 【说明】用户可以根据网络的实际状况增大或者降低 BFD 会话的本地检测倍数。比如对于比较稳定的链路，由于不需要频繁地检测链路状态，可以增大 BFD 会话的检测倍数。 缺省情况下，本地 BFD 检测倍数为 3，可用 **undo detect-multiplier** 命令恢复 BFD 会话的本地检测倍数为缺省值
10	**process-pst** 例如：[Huawei-bfd-session-1to4] **process-pst**	允许 BFD 会话状态改变时通告上层应用。如果允许 BFD 修改端口状态表 PST（Port State Table），当检测到 BFD 会话状态变为 Down 时，系统将更改 PST 中相应表项。 缺省情况下，静态 BFD 会话未使能通告联动检测业务，可用 **undo process-pst** 命令恢复缺省配置
11	**commit** 例如：[Huawei-bfd-session-1to4] **commit**	提交配置。无论改变任何 BFD 配置，必须执行 commit 命令，才能使配置生效。 【说明】BFD 会话建立需要满足一定的条件，包括绑定的接口状态是 Up、有去往 **peer-ip** 的可达路由。如果当前不满足会话建立条件，执行本命令后，系统将保留该会话的配置表项，但会话表项不能建立。但系统定期扫描已经提交但尚未建立会话的 BFD 配置表项，如果满足条件，则建立会话。 系统所允许建立的 BFD 会话有数量限制。当已经建立的 BFD 会话数达到上限时，如果对新的 BFD 会话执行本命令，系统将产生日志信息，提示无法创建会话，同时发送 Trap 消息

2. 配置出节点 BFD 参数

如果本端配置采用静态 LSP BFD 检测，对端所采用的 BFD 检测方式可以是多种方式，如是静态或动态 LSP BFD 检测、IP 链路 BFD 检测、TE 隧道 BFD 检测等。当然，后面章节将要介绍的动态 LSP BFD 检测、TE 隧道 BFD 检测也相同。

出节点可配置的 BFD 参数包括：所绑定的对端 IP 地址、本地标识符、远端标识符、本

地发送 BFD 报文的时间间隔、本地接收 BFD 报文的时间间隔和本地 BFD 检测倍数，这些将会影响会话的建立。用户可以根据网络的实际状况调整本地检测时间。对于不太稳定的链路，如果本地检测时间较短，则 BFD 会话可能会发生震荡，这时可以选择延长本地检测时间。

出节点的 BFD 参数配置步骤见表 2-3，与入节点的 BFD 会话配置方法基本一样，只不过在创建 BFD 会话时可根据反向通道的不同类型，选择不同的配置命令。为了保证 BFD 报文往返路径一致，一般情况下反向通道优先选用 LSP 或者 TE 隧道。

表 2-3　**出节点 BFD 参数的配置步骤**

步骤	命令	说明
1	**system-view** 例如：<Huawei> **system-view**	进入系统视图
2	**bfd** 例如：[Huawei] **bfd**	对本节点使能全局 BFD 能力并进入 BFD 全局视图。 缺省情况下，全局 BFD 功能未使能，可用 **undo bfd** 命令全局去使能 BFD 功能，此时如果已经配置了 BFD 会话信息，则所有的 BFD 会话都会被删除
3	**quit** 例如：[Huawei-bfd] **quit**	返回系统视图
4	**bfd** *cfg-name* **bind peer-ip** *peer-ip* [**vpn-instance** *vpn-instance-name*] [**interface** *interface-type interface-number*] [**source-ip** *source-ip*] 例如：[Huawei] **bfd** atoc **bind peer-ip** 10.10.20.2	（四选一）当反向通道是 IP 链路时创建 BFD 会话。在创建 BFD 会话时，单跳检测必须绑定对端 IP 地址和本端相应接口，多跳检测只需绑定对端 IP 地址。命令中的参数说明如下。 • *cfg-name*：指定 BFD 配置名，字符串形式，不支持空格，不区分大小写，长度范围是 1～15。当输入的字符串两端使用双引号时，可在字符串中输入空格。 • **peer-ip** *ip-address*：指定 BFD 会话绑定的对端 IP 地址。**如果只指定对端 IP 地址，则表示检测多跳链路。** • **vpn-instance** *vpn-name*：可选参数，指定对端 BFD 会话绑定的 VPN 实例名称，必须是已创建的 VPN 实例。如果不指定 VPN 实例，则认为对端地址是公网地址。如果同时指定了对端 IP 地址和 VPN 实例，则表示检测 VPN 路由的多跳链路。 • **interface** *interface-type interface-number*：可选参数，指定绑定 BFD 会话的接口。**如果同时指定了对端 IP 地址和本端接口，表示检测单跳链路**，即检测以该接口为出接口、以 peer-ip 为下一跳地址的一条固定路由；如果同时指定了对端 IP 地址、VPN 实例和本端接口，表示检测 VPN 路由的单跳链路。 • **source-ip** *ip-address*：可选参数，指定 BFD 报文携带的源 IP 地址。通常情况下，不需要配置该参数。在 BFD 会话协商阶段，如果不配置该参数，则系统将在本地路由表中查找去往对端 IP 地址的出接口，将该出接口的 IP 地址作为本端发送 BFD 报文的源 IP 地址。在 BFD 会话检测链路阶段，如果不配置该参数，则系统会将 BFD 报文的源 IP 地址设置为一个固定的值。 缺省情况下，没有创建 BFD 会话，可用 **undo bfd session-name** 命令删除指定的 BFD 会话，同时取消 BFD 会话的绑定信息

（续表）

步骤	命令	说明
4	**bfd** *cfg-name* **bind static-lsp** *lsp-name* 例如：[Huawei] **bfd** 1to4 **bind static-lsp** 1to4	（四选一）当反向通道是静态 LSP 时创建静态 LSP 的 BFD 会话，参数说明参见表 2-2 中的第 4 步
	bfd *cfg-name* **bind ldp-lsp peer-ip** *ip-address* **nexthop** *ip-address* [**interface** *interface-type interface-number*] 例如：Huawei] **bfd** 1to4 **bind ldp-lsp peer-ip** 4.4.4.4 **nexthop** 1.1.1.1 **interface** gigabitethernet 1/0/0	（四选一）当反向通道是动态 LSP 时创建 LDP LSP 的 BFD 会话。命令中的参数说明如下。 • *cfg-name*：指定 BFD 会话名称，字符串形式，不支持空格，不区分大小写，长度范围是 1～15。当输入的字符串两端使用双引号时，可在字符串中输入空格。 • **peer-ip** *ip-address*：指定 BFD 会话绑定动态 LSP 的目的端 IP 地址。 • **nexthop** *ip-address*：指定被检测 LSP 的下一跳 IP 地址。 • **interface** *interface-type interface-number* ：可选参数，指定 BFD 绑定的出接口。 缺省情况下，没有创建检测 LDP LSP 的 BFD 会话，可用 **undo bfd** *cfg-name* 命令删除指定的 BFD 会话
	bfd *cfg-name* **bind mpls-te interface tunnel** *interface-number* [**te-lsp** [**backup**]] 例如：[Huawei] **bfd** 1to4rsvp **bind mpls-te interface** Tunnel 0/0/1 **te-lsp**	（四选一）当反向通道是 TE 隧道时创建 BFD 会话或与 TE 隧道绑定的主用或备用 LSP。命令中的参数说明如下。 • *cfg-name*：指定创建 BFD 会话名称，字符串形式，不支持空格，不区分大小写，长度范围是 1～15。当输入的字符串两端使用双引号时，可在字符串中输入空格。 • **interface tunnel** *interface-number*：指定 BFD 会话绑定的 Tunnel 接口编号。 • **te-lsp** [**backup**]：**可选项**，指定 BFD 检测与 Tunnel 隧道绑定的 LSP。其中：未选择 **backup** 可选项时，指定 BFD 检测与 Tunnel 隧道绑定的主 LSP；选择了 **backup** 可选项时，指定 BFD 检测与 Tunnel 隧道绑定的备用LSP。BFD检测与Tunnel绑定的主用或备用LSP时，如果 LSP 的状态为 Down，则不能建立 BFD 会话。 BFD 检测 TE 隧道时，如果 TE 隧道的状态为 Down，则能够创建 BFD 会话，但 BFD 会话不能 Up。一个 TE 隧道可能有多个 LSP，当 BFD 检测 TE 隧道时，只有全部 LSP 都出现故障，BFD 会话的状态才为 Down。 缺省情况下，Tunnel 隧道没有使用 BFD 检测，可用 **undo bfd** *cfg-name* 命令删除指定的 BFD 会话
5	**discriminator local** *discr-value* 例如：[Huawei-bfd-session-1to4] **discriminator local** 10	配置本地标识符，参见表 2-2 中的第 5 步
6	**discriminator remote** *discr-value* 例如：[Huawei-bfd-session-1to4] **discriminator remote** 20	配置远端标识符，参见表 2-2 中的第 6 步
7	**min-tx-interval** *interval* 例如：[Huawei-bfd-session-1to4] **min-tx-interval** 300	（可选）调整本地发送 BFD 报文的时间间隔，参见表 2-2 中的第 7 步

（续表）

步骤	命令	说明
8	**min-rx-interval** *interval* 例如：[Huawei-bfd-session-1to4] **min-rx-interval** 600	（可选）调整本地接收 BFD 报文的时间间隔，参见表 2-2 中的第 8 步
9	**detect-multiplier** *multiplier* 例如：[Huawei-bfd-session-1to4] **detect-multiplier** 5	（可选）调整本地 BFD 检测倍数，参见表 2-2 中的第 9 步
10	**process-pst** 例如：[Huawei-bfd-session-1to4] **process-pst**	（可选）允许 BFD 会话状态改变时通告上层应用，参见表 2-2 中的第 10 步
11	**commit** 例如：[Huawei-bfd-session-1to4] **commit**	提交配置，参见表 2-2 中的第 11 步

2.1.3　检测静态 LSP 的连通性

在 MPLS 中，如果 LSP 转发数据失败，负责建立 LSP 的 MPLS 控制平面将无法检测到这种错误，这会给网络维护带来困难。我们已在第 1 章 1.3 节介绍过，MPLS Ping 主要用于检查 LSP 的连通性，MPLS Traceroute 在检查 LSP 的连通性的同时，还可以分析网络什么位置发生了故障。可以在任意视图下进行 MPLS Ping/Traceroute 测试，但 MPLS Ping/Traceroute 不支持分片报文，即不会对发送的请求和响应报文进行分片。

静态 LSP 连通性检测配置和操作步骤见表 2-4。

表 2-4　静态 LSP 连通性检测配置和操作步骤

步骤	命令	说明
1	**system-view** 例如：<Huawei> **system-view**	进入系统视图
2	**lspv mpls-lsp-ping echo enable** 例如：[Huawei] **undo lspv mpls-lsp-ping echo enable**	使能对 MPLS echo request 报文的响应功能。 缺省情况下，系统对 MPLS echo request 报文的响应功能是使能的，可用 **undo lspv mpls-lsp-ping echo enable** 命令关闭对 MPLS echo request 报文的响应功能，但这样会导致此设备不会对下面将要介绍的 **ping lsp** 和 **tracert lsp** 命令响应，造成目的地址为此设备的命令结果显示为超时
3	**lspv packet-filter** *acl-number* 例如：[Huawei] **lspv packet-filter** 2100	（可选）使能对 MPLS echo request 报文的源地址过滤功能，过滤规则在 ACL（可以是基本 ACL，也可以是高级 ACL）中指定。 如果使能了对 MPLS echo request 报文的源地址过滤功能，当收到 MPLS echo request 报文时，设备会使用指定的 ACL 对报文的源 IP 进行检查。ACL 条件允许的报文被继续处理；ACL 条件不允许的报文被丢弃。 缺省情况下，系统对 MPLS echo request 报文的源地址过滤功能是关闭的，可用 **undo lspv packet-filter** 命令关闭对 MPLS echo request 报文的源地址过滤功能

（续表）

步骤	命令	说明
4	**ping lsp** [**-a** *source-ip* | **-c** *count* | **-exp** *exp-value* | **-h** *ttl-value* | **-m** *interval* | **-r** *reply-mode* | **-s** *packet-size* | **-t** *time-out* | **-v**] * **ip** *destination-address mask-length* [*ip-address*] [**nexthop** *nexthop-address* | **draft6**] 例如：[Huawei] **ping lsp -c** 10 **-s** 200 **ip** 4.4.4.9 32	MPLS ping 测试可以在所有视图下执行。 为了防止消息到达 Egress 节点后又被转发给其他节点，MPLS Echo-Request 消息的 IP 报头中目的地址设置为 127.0.0.1/8（本机环回地址），IP 报头中的 TTL 值=1 命令中的参数和选项说明如下。 ● **-a** *source-ip*：可多选参数，指定发送 MPLS ECHO-REQUEST 报文的源 IP 地址。如果不指定源 IPv4 地址，将采用出接口的 IP 地址作为 MPLS ECHO-REQUEST 报文发送的源地址。 ● **-c** *count*：可多选参数，指定发送 MPLS ECHO-REQUEST 报文次数，整数形式，取值范围是 1～4294967295。缺省值是 5。 ● **-exp** *exp-value*：可多选参数，指定发送的 MPLS ECHO-REQUEST 请求报文的 EXP（MPLS 优先级）值，整数形式，取值范围是 0～7。缺省值是 0。 ● **-h** *ttl-value*：可多选参数，指定 TTL 的值，整数形式，取值范围是 1～255。缺省值是 64。 【说明】**ping lsp** 命令每发送一个 MPLS ECHO-REQUEST 报文，顺序号就加 1，顺序号从 1 开始，默认情况一共发送 5 个 MPLS ECHO-REQUEST 报文，也可以通过命令行参数 ttl-value 设置发送 MPLS ECHO-REQUEST 报文的个数。如果对端可达，则在对端会相应回应 5 个和请求端同样序号的 MPLS ECHO-REPLY 报文。报文在转发过程中，如果 TTL 字段的值减为 0，报文到达的路由器就会向源端发送超时报文，表明远程设备不可达。 ● **-s** *packet-size*：可多选参数，指定 MPLS ECHO-REQUEST 报文的净荷报文长度，即不包括 IP 头和 UDP 头的报文长度，整数形式，取值范围是 65～8100，单位是字节。缺省值是 100。配置的值要小于出接口的 MTU 值。 ● **-t** *time-out*：可多选参数，指定发送完 MPLS ECHO-REQUEST 后，等待 MPLS ECHO-REPLY 的超时时间，整数形式，取值范围是 0～65535，单位是毫秒。缺省值是 2000 毫秒。 ● **-v**：可多选选项，指定显示接收到的非本用户的 ICMP ECHO-RESPONSE 的 ICMP 报文，如果不指定，系统只显示本用户收到 ICMP ECHO-RESPONSE 报文。 ● *destination-address mask-length*：指定目的对端的 IPv4 地址和掩码长度。 ● *ip-address*：可选参数，指定在 MPLS ECHO-REQUEST 报文 IP 报头中封装的目的地址。缺省情况下，MPLS ECHO-REQUEST 报文 IP 头中的目的地址是 127.0.0.1。 ● **draft6**：可选项，按 draft-ietf-mpls-lsp-ping-06 实现。默认按 RFC4379 实现

（续表）

步骤	命令	说明
4	**tracert lsp** [**-a** *source-ip* | **-exp** *exp-value* | **-h** *ttl-value* | **-r** *reply-mode* | **-t** *time-out* | **-v**] * **ip** *destination-address mask- length* [*ip-address*] [**nexthop** *nexthop-address* | **draft6**] 例如：[Huawei] **tracert lsp ip** 8.4.4.9 32	MPLS Traceroute 测试可以在所有视图下执行。**tracert lsp** 命令的执行过程如下。 • 发送一个 TTL 为 1 的数据包，TTL 超时，第一跳发送回一个 MPLS ECHO-REPLY 报文。 • 发送一个 TTL 为 2 的数据包，TTL 超时，第二跳发送回一个 MPLS ECHO-REPLY 报文。 • 发送一个 TTL 为 3 的数据包，TTL 超时，第三跳发送回一个 MPLS ECHO-REPLY 报文。 上述过程不断进行，直到到达目的地。为了防止消息到达 Egress 节点后又被转发给其他节点，MPLS ECHO-REQUEST 消息的 IP 头中目的地址设置为环回地址，前缀为 127.0.0.1/8。 本命令中的许多参数与上面的 **ping lsp** 命令中的参数及功能说明完全相同，下面仅介绍在上一命令中没有或者功能说明不一样的参数或选项说明。 • **-r** *reply-mode*：可多选参数，指定对端回送 MPLS echo reply 报文的模式，整数形式，取值范围为 1～4。缺省值是 2。1 为不应答，2 为通过 IPv4 UDP 报文应答，3 为通过带 Router alert 的 IPv4 UDP 报文应答，4 为通过应用平面的控制通道应答。如果设置 reply-mode 为 1，则进行单向测试，测试发起端显示超时表明测试成功，否则会提示 LSP 不存在。 • **-v**：可多选选项，指定显示 ICMP Time Exceeded 报文带回的 MPLS 标签信息。该参数在 PE 上发起 tracert 需要显示公网标签时使用。 • **nexthop** *nexthop-address*：可选参数，指定下一跳 IP 地址

2.1.4　静态 LSP 及 BFD 检测维护与管理

已经完成静态 LSP 和 BFD 检测功能的配置后，可在任意视图下通过以下 **display** 命令查看相关配置或统计信息，以验证配置结果。

■ **display default-parameter mpls management**：查看 MPLS 管理的缺省配置。

■ **display mpls interface** [*interface-type interface-number*] [**verbose**]：查看所有或指定接口使能 MPLS 的情况。

■ **display mpls static-lsp** [*lsp-name*] [{ **include** | **exclude** } *ip-address mask-length*] [**verbose**]：查看指定或所有静态 LSP 的配置信息。

■ **display mpls label static available** [[**label-from** *label-index*] **label-number** *label-number*]：查看当前静态业务可以使用的 LSP 标签（当前，在取值范围中没有分配的标签）。

■ **display bfd configuration** { **all** | **static** } [**for-lsp**]：查看所有或静态的 LSP BFD 配置信息。

■ **display bfd session** { **all** | **static** } [**for-lsp**]：查看所有或静态的 LSP BFD 会话信息。

■ **display bfd statistics session** { **all** | **static** } [**for-ip** | **for-lsp**]，查看所有或静态的 IP 或 LSP 的 BFD 会话统计信息。

■ **display mpls static-lsp** [*lsp-name*] [{ **include** | **exclude** } *ip-address mask-length*] [**verbose**]：查看所有或指定 FEC 关联的静态 LSP 的状态。

■ **display lspv statistics**：查看 LSPV 的统计结果信息。

■ **display lspv configuration**：查看 LSPV 当前的配置信息。

2.1.5 AR 路由器静态 LSP 配置示例

如图 2-1 所示，LSR_1、LSR_2、LSR_3 为某 MPLS 骨干网设备。现要求在骨干网上创建稳定的公网隧道来承载 L2VPN 或 L3VPN 业务。

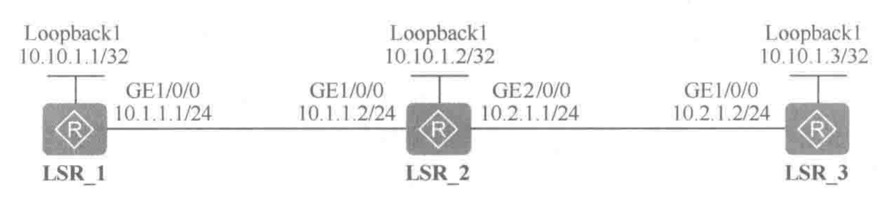


图 2-1 AR 路由器静态 LSP 配置示例的拓扑结构

1. 基本配置思路分析

因为本示例的拓扑结构简单且稳定，所以可采用静态 LSP 配置方式。同时，因为 LSP 是单向的，所以如果要实现各设备所连网络互通，则需要配置两条静态 LSP：一条是由 LSR_1 到 LSR_3 的 LSP(假设名称为 LSP1)，此时 LSR_1 为 Ingress，LSR_2 为 Transit，LSR_3 为 Egress；另一条是由 LSR_3 到 LSR_1 的 LSP（假设名称为 LSP2），此时 LSR_3 为 Ingress，LSR_2 为 Transit，LSR_1 为 Egress。

根据 1.2.1 小节介绍的配置步骤，再结合本示例实际，可得出本示例如下的配置思路。

（1）在各 LSR 上配置 OSPF 协议（当然也可以是静态路由或其他 IGP），实现骨干网的 IP 连通性，这是前提。因为在 LSP 的配置中需要利用 IP 路由进行 FEC 划分，也需要利用 IP 路由来确保下一跳可达。

（2）在 LSR 上配置 LSR ID，使能全局和公网接口的 MPLS 能力，这是实现在骨干网上创建公网隧道的前提。

（3）在两条 LSP 的 Ingress 上配置目的地址、下一跳和出标签的值；在 Transit 上配置入接口、与上游节点出标签相同的入标签值、对应的下一跳 IP 地址和出标签的值；在 Egress 上配置入接口、与上游节点出标签相同的入标签值。

2. 具体配置步骤

（1）配置 LSR 各接口（包括 Loopback 接口）的 IP 地址。

\# LSR_1 上的配置。

```
<Huawei> system-view
[Huawei] sysname LSR_1
[LSR_1] interface loopback 1
[LSR_1-LoopBack1] ip address 10.10.1.1 32
[LSR_1-LoopBack1] quit
[LSR_1] interface gigabitethernet 1/0/0
[LSR_1-GigabitEthernet1/0/0] ip address 10.1.1.1 24
[LSR_1-GigabitEthernet1/0/0] quit
```

#　LSR_2 上的配置。

```
<Huawei> system-view
[Huawei] sysname LSR_2
[LSR_2] interface loopback 1
[LSR_2-LoopBack1] ip address 10.10.1.2 32
[LSR_2-LoopBack1] quit
[LSR_2] interface gigabitethernet 1/0/0
[LSR_2-GigabitEthernet1/0/0] ip address 10.1.1.2 24
[LSR_2-GigabitEthernet1/0/0] quit
[LSR_2] interface gigabitethernet 2/0/0
[LSR_2-GigabitEthernet2/0/0] ip address 10.2.1.1 24
[LSR_2-GigabitEthernet2/0/0] quit
```

#　LSR_3 上的配置。

```
<Huawei> system-view
[Huawei] sysname LSR_3
[LSR_3] interface loopback 1
[LSR_3-LoopBack1] ip address 10.10.1.3 32
[LSR_3-LoopBack1] quit
[LSR_3] interface gigabitethernet 1/0/0
[LSR_3-GigabitEthernet1/0/0] ip address 10.2.1.2 24
[LSR_3-GigabitEthernet1/0/0] quit
```

（2）配置 OSPF 路由，把各接口所连网段的路由，都加入到 OSP 路由进程 1，区域 0 中（单区域 OSPF 网络时区域 ID 任意）。

#　LSR_1 上的配置。

```
[LSR_1] ospf 1
[LSR_1-ospf-1] area 0
[LSR_1-ospf-2-area-0.0.0.0] network 10.10.1.1 0.0.0.0
[LSR_1-ospf-2-area-0.0.0.0] network 10.1.1.0 0.0.0.255
[LSR_1-ospf-2-area-0.0.0.0] quit
[LSR_1-ospf-1] quit
```

#　LSR_2 上的配置。

```
[LSR_2] ospf 1
[LSR_2-ospf-1] area 0
[LSR_2-ospf-2-area-0.0.0.0] network 10.10.1.2 0.0.0.0
[LSR_2-ospf-2-area-0.0.0.0] network 10.1.1.0 0.0.0.255
[LSR_2-ospf-2-area-0.0.0.0] network 10.2.1.0 0.0.0.255
[LSR_2-ospf-2-area-0.0.0.0] quit
[LSR_2-ospf-1] quit
```

#　LSR_3 上的配置。

```
[LSR_3] ospf 1
[LSR_3-ospf-1] area 0
[LSR_3-ospf-2-area-0.0.0.0] network 10.10.1.3 0.0.0.0
[LSR_3-ospf-2-area-0.0.0.0] network 10.2.1.0 0.0.0.255
[LSR_3-ospf-2-area-0.0.0.0] quit
[LSR_3-ospf-1] quit
```

配置好 OSPF 路由后，在各节点上执行 **display ip routing-table** 命令，可以看到相互之间都学到了彼此的路由。

（3）配置 MPLS 基本功能。LSR ID 以各自的 Loopback1 接口 IP 地址进行配置，在全局以及 LSP 路径上各 LSR 间互联的接口上使能 MPLS 功能。

#　LSR_1 上的配置。

```
[LSR_1] mpls lsr-id 10.10.1.1
[LSR_1] mpls
[LSR_1-mpls] quit
[LSR_1] interface gigabitethernet 1/0/0
[LSR_1-GigabitEthernet1/0/0] mpls
[LSR_1-GigabitEthernet1/0/0] quit
```

LSR_2 上的配置。

```
[LSR_2] mpls lsr-id 10.10.1.2
[LSR_2] mpls
[LSR_2-mpls] quit
[LSR_2] interface gigabitethernet 1/0/0
[LSR_2-GigabitEthernet1/0/0] mpls
[LSR_2-GigabitEthernet1/0/0] quit
[LSR_2] interface gigabitethernet 2/0/0
[LSR_2-GigabitEthernet2/0/0] mpls
[LSR_2-GigabitEthernet2/0/0] quit
```

LSR_3 上的配置。

```
[LSR_3] mpls lsr-id 10.10.1.3
[LSR_3] mpls
[LSR_3-mpls] quit
[LSR_3] interface gigabitethernet 1/0/0
[LSR_3-GigabitEthernet1/0/0] mpls
[LSR_3-GigabitEthernet1/0/0] quit
```

（4）创建静态 LSP。因为这里涉及两个方向的两条 LSP，LSR_1 和 LSR_3 在不同 LSP 中的角色不一样。

■ 创建从 LSR_1 到 LSR_3 的静态 LSP1。

Ingress LSR_1 上的配置。配置目的 IP 地址（LSR_3 的 Loopback1 接口 IP 地址）、下一跳和出标签（假设为 20）。

```
[LSR_1] static-lsp ingress LSP1 destination 10.10.1.3 32 nexthop 10.1.1.2 out-label 20
```

Transit LSR_2 上的配置。配置入接口、入标签（20，要与 LSR_1 的出标签一致）、下一跳和出标签（假设为 40）。

```
[LSR_2] static-lsp transit LSP1 incoming-interface gigabitethernet 1/0/0 in-label 20 nexthop 10.2.1.2 out-label 40
```

Egress LSR_3 上的配置。配置入接口和入标签（40，要与 LSR_2 的出标签一致）。

```
[LSR_3] static-lsp egress LSP1 incoming-interface gigabitethernet 1/0/0 in-label 40
```

说明

本示例中并没有说明 LSR_1 和 LSR_3 后面各自连接的是否是 IP 网络，所以在出节点的入标签（也对应倒数第二节点的出标签）配置上没有配置强制弹出标签所用的特殊标签 3。

配置完成后，可在各节点上执行 **display mpls static-lsp** 命令查看静态 LSP 的状态。以下是在 LSR_1 上执行本命令的输出示例，从中可以看出有一条名为 LSP1 的 LSP，关联的 FEC 是 10.0.1.3/32，没有入标签（在 Ingress 节点上无需配置入标签），出标签为 20，出接口为 GE1/0/0，状态为 Up。

```
[LSR_1] display mpls static-lsp
TOTAL          : 1        STATIC LSP(S)
UP             : 1        STATIC LSP(S)
DOWN           : 0        STATIC LSP(S)
Name           FEC              I/O Label    I/O If          Status
LSP1           10.10.1.3/32     NULL/20      -/GE1/0/0       Up
```

■ 创建从LSR_3到LSR_1的静态LSP2。

Ingress LSR_3上的配置。配置目的IP地址（LSR_1的Loopback1接口IP地址）、下一跳和出标签（假设为30）。

```
[LSR_3] static-lsp ingress LSP2 destination 10.10.1.1 32 nexthop 10.2.1.1 out-label 30
```

Transit LSR_2上的配置。配置入接口、入标签（30，要与LSR_1的出标签一致）、下一跳和出标签（假设为60）。

```
[LSR_2] static-lsp transit LSP2 incoming-interface gigabitethernet 2/0/0 in-label 30 nexthop 10.1.1.1 out-label 60
```

Egress LSR_1上的配置。配置入接口和入标签（60，要与LSR_2的出标签一致）。

```
[LSR_1] static-lsp egress LSP2 incoming-interface gigabitethernet 1/0/0 in-label 60
```

3. 实验结果验证

两个方向的LSP都配置好后，可在各节点上执行 **display mpls static-lsp** 或 **display mpls static-lsp verbose** 命令查看静态LSP的状态或详细信息。以下是在LSR_3上执行这两条命令的输出示例，LSP的状态均为Up。

```
[LSR_3] display mpls static-lsp
TOTAL          : 2     STATIC LSP(S)   #---显示有两条静态LSP
Up             : 2     STATIC LSP(S)   #---这两条LSP的状态都是Up
DOWN           : 0     STATIC LSP(S)
Name       FEC                 I/O Label   I/O If              Status
LSP1       -/-                 40/NULL     GE1/0/0/-           Up
LSP2       10.10.1.1/32        NULL/30     -/GE1/0/0           Up

[LSR_3] display mpls static-lsp verbose     #---执行这行命令会显示两条LSP的详细配置信息
No              : 1
LSP-Name        : LSP1
LSR-Type        : Egress
FEC             : -/-
In-Label        : 40
Out-Label       : NULL
In-Interface    : GigabitEthernet1/0/0
Out-Interface   : -
NextHop         : -
Static-Lsp Type: Normal
Lsp Status      : Up

No              : 2
LSP-Name        : LSP2
LSR-Type        : Ingress
FEC             : 10.10.1.1/32
In-Label        : NULL
Out-Label       : 30
In-Interface    : -
Out-Interface   : GigabitEthernet1/0/0
NextHop         : 10.2.1.1
Static-Lsp Type: Normal
Lsp Status      : Up
```

此时，在LSR_3上执行 **ping lsp ip** 10.10.1.1 32命令，Ping到达LSR_1 Loopback1接口IP地址的LSP是通的，测试结果如下所示。同样在LSR_1上执行 **ping lsp ip** 10.10.1.3 32命令，Ping到达LSR_3 Loopback1接口IP地址的LSP也是通的。证明前面的静态LSP

配置是正确的，LSP 已成功建立。还可执行普通的 **Ping** 命令来测试各节点间的路由互通。

```
<LSR_3>ping lsp ip 10.10.1.1 32
  LSP ping FEC: IPV4 PREFIX 10.10.1.1/32/ : 100  data bytes, press CTRL_C to break
    Reply from 10.10.1.1: bytes=100 Sequence=1 time=90 ms
    Reply from 10.10.1.1: bytes=100 Sequence=2 time=90 ms
    Reply from 10.10.1.1: bytes=100 Sequence=3 time=90 ms
    Reply from 10.10.1.1: bytes=100 Sequence=4 time=70 ms
    Reply from 10.10.1.1: bytes=100 Sequence=5 time=60 ms

  --- FEC: IPV4 PREFIX 10.10.1.1/32 ping statistics ---
    5 packet(s) transmitted
    5 packet(s) received
    0.00% packet loss
    round-trip min/avg/max = 60/80/90 ms
```

下面再来验证本章前面所说的，在纯 LSP 隧道（非 MPLS VPN 应用）中，MPLS 最多只会带一层 MPLS 标签。先对 LSR_2 的 G0/0/0 端口在 LSR_1 ping LSR_3 时进行抓包，随便找一个抓取的 MPLS 报文，均发现其中只有一层 MPLS 标签（图中的"Mutiprotocol Label Switching Echo"是回显信息，是该 MPLS 报文的真正数据部分）。图 2-2 中所示的第二个 MPLS 报文中上面所携带的标签为 20。

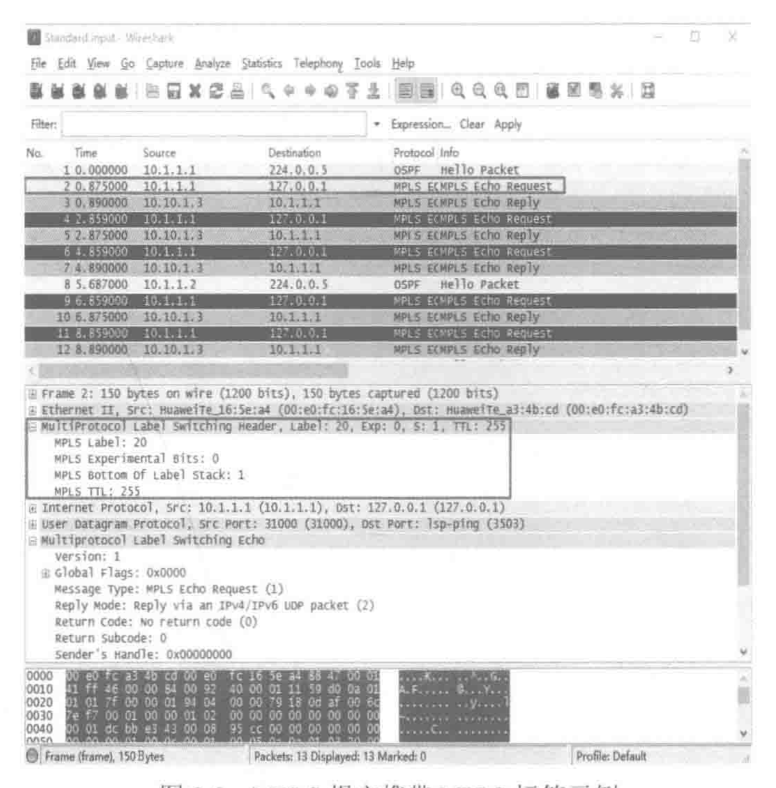


图 2-2　MPLS 报文携带 MPLS 标签示例

2.1.6　S 交换机静态 LSP 配置示例

因为在 S 系列交换机目前的 VRP 系统版本中，物理以太网接口暂不能直接配置 IP 地址，所以在 MPLS/IP 骨干网三层互通的配置方面与 AR 系列路由的配置有些不同（S 系列交换机要通过 VLANIF 接口来配置三层互通，并要在 VLANIF 接口使能 MPLS 功

能），故在此单独举例介绍。

如图 2-3 所示，网络拓扑结构简单并且稳定，LSR_1、LSR_2、LSR_3 为 MPLS 骨干网设备。要求在骨干网上创建稳定的公网隧道来承载 L2VPN 或 L3VPN 业务。

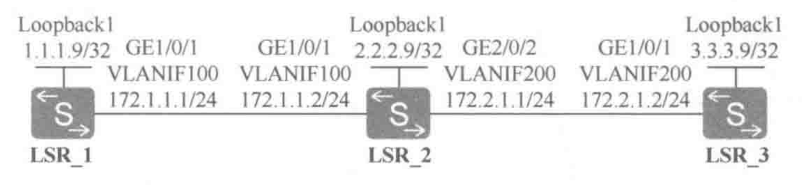


图 2-3　S 交换机静态 LSP 配置示例的拓扑结构

1. 基本配置思路分析

本示例其实与 2.1.5 小节的示例是一样的，不同的只是这里的设备全是 S 系列交换机，涉及通过 VLANIF 接口来实现邻居间动态路由的学习。这里关键是相邻设备间链路两端的二层端口配置。

本示例也是要通过配置静态 LSP 来实现在 LSR_1 和 LSR_3 之间建立公网 LSP 隧道。按要求需配置两条静态 LSP：LSR_1 到 LSR_3 的路径为 LSP1，LSR_1 为 Ingress，LSR_2 为 Transit，LSR_3 为 Egress；LSR_3 到 LSR_1 的路径为 LSP2，LSR_3 为 Ingress，LSR_2 为 Transit，LSR_1 为 Egress。

本示例的基本配置思路如下。

（1）在各 LSR 的各接口上配置 VLAN 和接口 IP 地址。

（2）在各 LSR 上配置 OSPF，实现骨干网的 IP 连通性。

（3）在 LSR 上配置 MPLS 功能，这是实现在骨干网上创建公网隧道的前提。

（4）在两条 LSP 的 Ingress 上配置目的地址、下一跳和出标签的值；在 Transit 上配置入接口、与上游节点出标签相同的入标签值、对应的下一跳 IP 地址和出标签值；在 Egress 上配置入接口、与上游节点出标签相同的入标签的值。

2. 具体配置步骤

（1）在各交换机上创建 VLAN、VLANIF 接口，配置 VLANIF 接口的 IP 地址，并将相应的物理接口加入到 VLAN，以激活对应的 VLANIF 接口。

LSR_1 上的配置。

```
<HUAWEI> system-view
[HUAWEI] sysname LSR_1
[LSR_1] interface loopback 1
[LSR_1-LoopBack1] ip address 1.1.1.9 32
[LSR_1-LoopBack1] quit
[LSR_1] vlan batch 100
[LSR_1] interface vlanif 100
[LSR_1-Vlanif100] ip address 172.1.1.1 24
[LSR_1-Vlanif100] quit
[LSR_1] interface gigabitethernet 1/0/1
[LSR_1-GigabitEthernet1/0/1] port link-type trunk
[LSR_1-GigabitEthernet1/0/1] port trunk allow-pass vlan 100
[LSR_1-GigabitEthernet1/0/1] quit
```

LSR_2 上的配置。

```
<HUAWEI> system-view
[HUAWEI] sysname LSR_2
```

```
[LSR_2] interface loopback 1
[LSR_2-LoopBack1] ip address 2.2.2.9 32
[LSR_2-LoopBack1] quit
[LSR_2] vlan batch 100
[LSR_2] interface vlanif 100
[LSR_2-Vlanif100] ip address 172.1.1.2 24
[LSR_2-Vlanif100] quit
[LSR_2] interface vlanif 200
[LSR_2-Vlanif200] ip address 172.2.1.1 24
[LSR_2-Vlanif200] quit
[LSR_2] interface gigabitethernet 1/0/1
[LSR_2-GigabitEthernet1/0/1] port link-type trunk
[LSR_2-GigabitEthernet1/0/1] port trunk allow-pass vlan 100
[LSR_2-GigabitEthernet1/0/1] quit
[LSR_2] interface gigabitethernet 2/0/2
[LSR_2-GigabitEthernet2/0/2] port link-type trunk
[LSR_2-GigabitEthernet2/0/2] port trunk allow-pass vlan 100
[LSR_2-GigabitEthernet2/0/2] quit
```

【经验提示】以上 LSR_1 中的 GE1/01 和 LSR_2 中的 GE1/0/1 接口不一定要配置为 Trunk 类型，且端口类型也不一定要相同，**只要能保证一端发送的 VLAN 数据帧对端端口可通过即可**，且其都可以成功激活对应的 VLANIF 接口。可根据实际情况配置。如对于 LSR_1 中的 GE1/01 和 LSR_2 中的 GE1/0/1 接口 VLAN 除了以上配置方法之外，下面再举两种配置方法（还有其他可选择的方法，如 Hybrid 和 Trunk 类型组合、Hybrid 和 Access 类型组合、都是 Hybrid 类型等）。

方法一：两端都采用 Access 类型，且都加入 VLAN 100 中。

■ LSR_1 的 GE1/0/1 配置。

```
[LSR_1] interface gigabitethernet 1/0/1
[LSR_1-GigabitEthernet1/0/1] port link-type access
[LSR_1-GigabitEthernet1/0/1] port default vlan 100
[LSR_1-GigabitEthernet1/0/1] quit
```

■ LSR_2 的 GE1/0/1 配置。

```
[LSR_2] interface gigabitethernet 1/0/1
[LSR_2-GigabitEthernet1/0/1] port link-type access
[LSR_2-GigabitEthernet1/0/1] port default vlan 100
[LSR_2-GigabitEthernet1/0/1] quit
```

方法二：一端为 Trunk 类型，另一端为 Access 类型。

如 LSR_1 的 GE1/01 为 Trunk 类型，LSR_2 的 GE1/01 为 Access 类型，但要修改 Trunk 端口的 PVID 值为 100，否则 LSR_1 的 GE1/01 端口发送的数据帧带的是缺省的 VLAN1 标签，而 LSR_2 的 GE1/01 以 Access 类型加入了 VLAN 100，是不能接收带 VLAN 1 标签的数据帧的。

■ LSR_1 的 GE1/0/1 配置。

```
[LSR_1] interface gigabitethernet 1/0/1
[LSR_1-GigabitEthernet1/0/1] port link-type trunk
[LSR_1-GigabitEthernet1/0/1] port trunk allow-pass vlan 100
[LSR_1-GigabitEthernet1/0/1] port trunk pvid vlan 100   #---必须要把 PVID 改为 VLAN 100，否则链路两端 LSR 建立不了 OSPF 邻居关系
[LSR_1-GigabitEthernet1/0/1] quit
```

■ LSR_2 的 GE1/0/1 配置。

```
[LSR_2] interface gigabitethernet 1/0/1
[LSR_2-GigabitEthernet1/0/1] port link-type access
[LSR_2-GigabitEthernet1/0/1] port default vlan 100
[LSR_2-GigabitEthernet1/0/1] quit
```

LSR_3 上的配置。

```
<HUAWEI> system-view
[HUAWEI] sysname LSR_3
[LSR_3] interface loopback 1
[LSR_3-LoopBack1] ip address 3.3.3.9 32
[LSR_3-LoopBack1] quit
[LSR_3] vlan batch 200
[LSR_3] interface vlanif 200
[LSR_3-Vlanif200] ip address 172.2.1.2 24
[LSR_3-Vlanif200] quit
[LSR_3] interface gigabitethernet 1/0/1
[LSR_3-GigabitEthernet1/0/1] port link-type trunk
[LSR_3-GigabitEthernet1/0/1] port trunk allow-pass vlan 200
[LSR_3-GigabitEthernet1/0/1] quit
```

（2）配置 OSPF 协议发布各节点接口所连网段和 LSR ID 的主机路由，都加入缺省的 OSPF 1 进程、区域 0 中。注意要同时发布各 LSR 作为 LSR ID 的 32 位掩码的 Loopback 接口地址路由。

LSR_1 上的配置。

```
[LSR_1] ospf 1
[LSR_1-ospf-1] area 0
[LSR_1-ospf-2-area-0.0.0.0] network 1.1.1.9 0.0.0.0
[LSR_1-ospf-2-area-0.0.0.0] network 172.1.1.0 0.0.0.255
[LSR_1-ospf-2-area-0.0.0.0] quit
[LSR_1-ospf-1] quit
```

LSR_2 上的配置。

```
[LSR_2] ospf 1
[LSR_2-ospf-1] area 0
[LSR_2-ospf-2-area-0.0.0.0] network 2.2.2.9 0.0.0.0
[LSR_2-ospf-2-area-0.0.0.0] network 172.1.1.0 0.0.0.255
[LSR_2-ospf-2-area-0.0.0.0] network 172.2.1.0 0.0.0.255
[LSR_2-ospf-2-area-0.0.0.0] quit
[LSR_2-ospf-1] quit
```

LSR_3 上的配置。

```
[LSR_3] ospf 1
[LSR_3-ospf-1] area 0
[LSR_3-ospf-2-area-0.0.0.0] network 3.3.3.9 0.0.0.0
[LSR_3-ospf-2-area-0.0.0.0] network 172.2.1.0 0.0.0.255
[LSR_3-ospf-2-area-0.0.0.0] quit
[LSR_3-ospf-1] quit
```

以上配置完成后，在各节点上执行 **display ip routing-table** 命令，可以看到相互之间都学到了彼此的路由。以下是在 LSR_1 上执行该命令的输出示例，从中可以看出 LSR_1 已学习了 LSR_2 和 LSR_3 上所有 OSPF 路由。

```
<LSR_1>display ospf routing

  OSPF Process 1 with Router ID 172.1.1.1
           Routing Tables
```

```
Routing for Network
Destination      Cost  Type      NextHop     AdvRouter    Area
1.1.1.9/32       0     Stub      1.1.1.9     172.1.1.1    0.0.0.0
172.1.1.0/24     1     Transit   172.1.1.1   172.1.1.1    0.0.0.0
2.2.2.9/32       1     Stub      172.1.1.2   172.1.1.2    0.0.0.0
3.3.3.9/32       2     Stub      172.1.1.2   172.2.1.2    0.0.0.0
172.2.1.0/24     2     Transit   172.1.1.2   172.2.1.2    0.0.0.0

Total Nets: 5
Intra Area: 5  Inter Area: 0  ASE: 0  NSSA: 0
```

（3）在各 LSR 上配置 MPLS 基本功能，包括 MPLS LSR ID 的配置，全局和在 VLANIF 接口上使能 MPLS 功能。

LSR_1 上的配置。

```
[LSR_1] mpls lsr-id 1.1.1.9
[LSR_1] mpls
[LSR_1-mpls] quit
[LSR_1] interface vlanif 100
[LSR_1-Vlanif100] mpls
[LSR_1-Vlanif100] quit
```

LSR_2 上的配置。

```
[LSR_2] mpls lsr-id 2.2.2.9
[LSR_2] mpls
[LSR_2-mpls] quit
[LSR_2] interface vlanif 100
[LSR_2-Vlanif100] mpls
[LSR_2-Vlanif100] quit
[LSR_2] interface vlanif 200
[LSR_2-Vlanif200] mpls
[LSR_2-Vlanif200] quit
```

LSR_3 上的配置。

```
[LSR_3] mpls lsr-id 3.3.3.9
[LSR_3] mpls
[LSR_3-mpls] quit
[LSR_3] interface vlanif 200
[LSR_3-Vlanif200] mpls
[LSR_3-Vlanif200] quit
```

（4）创建两条相反方向的静态 LSP。

■ 从 LSR_1 到 LSR_3 的静态 LSP，假设 LSP 名称为 LSP1。

Ingress LSR_1 上的配置。目的 IP 地址就是 LSR_3 的 Loopback1 接口的 IP 地址 3.3.3.9/32，假设所分配的出标签为 20。

```
[LSR_1] static-lsp ingress LSP1 destination 3.3.3.9 32 nexthop 172.1.1.2 out-label 20
```

Transit LSR_2 上的配置。所分配的入标签与 LSR_1 上为 LSP1 分配的出标签一致，为 20，出标签假设为 40。

```
[LSR_2] static-lsp transit LSP1 incoming-interface vlanif 100 in-label 20 nexthop 172.2.1.2 out-label 40
```

Egress LSR_3 上的配置。所分配的入标签与 LSR_2 上为 LSP1 分配的出标签一致，为 40。

```
[LSR_3] static-lsp egress LSP1 incoming-interface vlanif 200 in-label 40
```

以上配置完成后，可在各节点上用 **display mpls static-lsp** 命令查看静态 LSP 的状态。以下是在 LSR_1 上执行该命令的输出示例，从中可以看出 LSP1 已建立好，因为其状态

为 Up。

```
[LSR_1] display mpls static-lsp
TOTAL            : 1        STATIC LSP(S)
UP               : 1        STATIC LSP(S)
DOWN             : 0        STATIC LSP(S)
Name             FEC                 I/O Label   I/O If              Status
LSP1             3.3.3.9/32          NULL/20     -/Vlanif100         Up
```

■ 创建从 LSR_3 到 LSR_1 静态 LSP，假设 LSP 名称为 LSP2。

Ingress LSR_3 上的配置。目的 IP 地址就是 LSR_1 的 Loopback1 接口的 IP 地址 1.1.1.9/32，假设所分配的出标签为 30。

```
[LSR_3] static-lsp ingress LSP2 destination 1.1.1.9 32 nexthop 172.2.1.1 out-label 30
```

Transit LSR_2 上的配置。所分配的入标签与 LSR_3 上为 LSP2 分配的出标签一致，为 30，出标签假设为 60。

```
[LSR_2] static-lsp transit LSP2 incoming-interface vlanif 200 in-label 30 nexthop 172.1.1.1 out-label 60
```

Egress LSR_1 上的配置。所分配的入标签与 LSR_2 上为 LSP1 分配的出标签一致，为 60。

```
[LSR_1] static-lsp egress LSP2 incoming-interface vlanif 100 in-label 60
```

3. 配置结果验证

以上配置完成后，在各节点上用 **display mpls static-lsp** 或 **display mpls static-lsp verbose** 命令查看静态 LSP 的状态及其详细信息。以下是在 LSR_3 上执行这两条命令的输出示例。

```
[LSR_3] display mpls static-lsp
TOTAL            : 2        STATIC LSP(S)
Up               : 2        STATIC LSP(S)
DOWN             : 0        STATIC LSP(S)
Name             FEC                 I/O Label   I/O If              Status
LSP1             -/-                 40/NULL     Vlanif200/-         Up
LSP2             1.1.1.9/32          NULL/30     -/Vlanif200         Up

[LSR_3] display mpls static-lsp verbose
No                  : 1
LSP-Name            : LSP1
LSR-Type            : Egress
FEC                 : -/-
In-Label            : 40
Out-Label           : NULL
In-Interface        : Vlanif200
Out-Interface       : -
NextHop             : -
Static-Lsp Type: Normal
Lsp Status          : Up

No                  : 2
LSP-Name            : LSP2
LSR-Type            : Ingress
FEC                 : 1.1.1.9/32
In-Label            : NULL
Out-Label           : 30
In-Interface        : -
Out-Interface       : Vlanif200
```

```
NextHop            : 172.2.1.1
Static-Lsp Type: Normal
Lsp Status         : Up
```

此时 LSR_1 和 LSR_3 可以相互 Ping 通了，注意这里执行的是 MPLS 的 Ping 操作。以下是在 LSR_1 执行 **ping lsp ip** 3.3.3.9 32 命令，测试与 LSR_3 的通信的输出示例，结果显示是可以 Ping 通的。

```
<LSR_1>ping lsp ip 3.3.3.9 32
  LSP ping FEC: IPV4 PREFIX 3.3.3.9/32/ : 100   data bytes, press CTRL_C to break

    Reply from 3.3.3.9: bytes=100 Sequence=1 time=80 ms
    Reply from 3.3.3.9: bytes=100 Sequence=2 time=50 ms
    Reply from 3.3.3.9: bytes=100 Sequence=3 time=90 ms
    Reply from 3.3.3.9: bytes=100 Sequence=4 time=70 ms
    Reply from 3.3.3.9: bytes=100 Sequence=5 time=90 ms

  --- FEC: IPV4 PREFIX 3.3.3.9/32 ping statistics ---
    5 packet(s) transmitted
    5 packet(s) received
    0.00% packet loss
    round-trip min/avg/max = 50/76/90 ms
```

2.1.7 静态 BFD 检测静态 LSP 配置示例

如图 2-4 所示，PE、P 为 MPLS 骨干网设备，现决定采用静态 LSP（有主、备两条链路）来承载网络业务。目前，由于网络业务对实时性的要求越来越高，例如 VoIP、在线游戏、在线视频业务等，链路发生故障导致的数据丢失会对这些业务造成比较严重的影响，所以要求当主用 LSP 故障时，流量能够快速切换到备份 LSP，尽可能地避免流量的丢失。

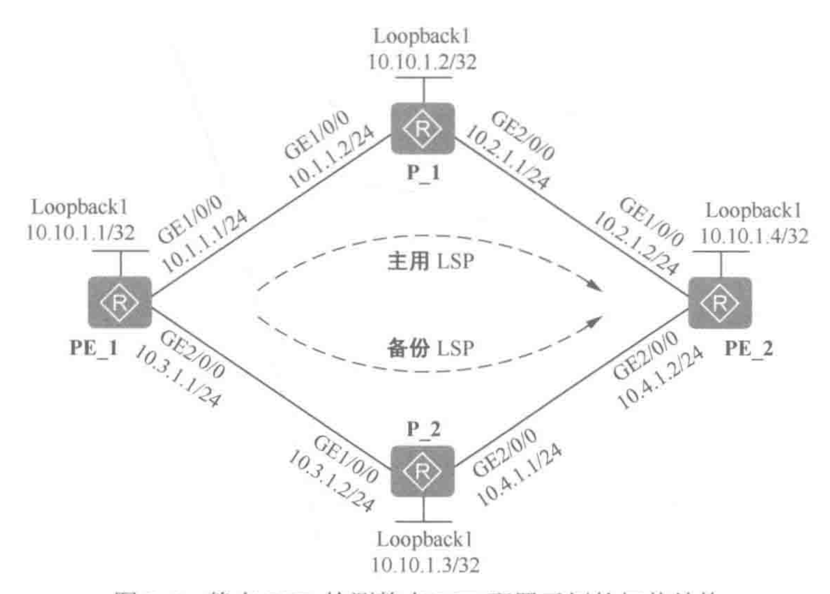


图 2-4 静态 BFD 检测静态 LSP 配置示例的拓扑结构

1. 基本配置思路分析

本示例中两 PE 设备间有主、备两条链路，要实现 MPLS 网络互通，则需要配置 4

条静态 LSP（每条链路各包括双向两条 LSP），假设主用链路中对应的两个方向的 LSP 名称分别为 LSP1 和 LSP2，备用链路中对应的两个方向的 LSP 名称分别为 LSP3、LSP4。如果要实现双向检测，主、备链路也各要配置双向两个静态 LSP BFD 会话。本示例中仅以主用链路上从 PE_1 到 PE_2 单向的静态 LSP BFD 检测为例进行介绍。

本示例的基本配置思路如下。

（1）在各 PE、P 上配置各接口（包括 Loopback 接口）的 IP 地址。

（2）在各 PE、P 上配置 OSPF，实现骨干网的 IP 连通性。并通过提高 PE_1 和 PE_2 节点的 GE2/0/0 接口的 OSPF 开销值，使得通过 P_2 的链路为备用链路。

（3）在各 PE、P 上配置两条链路、两个方向的静态 LSP，实现通过静态 LSP 承载网络业务的目的。

（4）在 PE_1 上配置通过主用链路到达 PE_2 的静态 LSP BFD 会话，实现对从 PE_1 到 PE_2 方向 LSP 的快速检测。如果同时要对其他 LSP 进行检测，配置方法类似。

2. 具体配置步骤

（1）配置各接口（包括 Loopback 接口）的 IP 地址。

PE_1 上的配置。

```
<Huawei> system-view
[Huawei] sysname PE_1
[PE_1] interface loopback 1
[PE_1-LoopBack1] ip address 10.10.1.1 32
[PE_1-LoopBack1] quit
[PE_1] interface gigabitethernet 1/0/0
[PE_1-GigabitEthernet1/0/0] ip address 10.1.1.1 24
[PE_1-GigabitEthernet1/0/0] quit
[PE_1] interface gigabitethernet 2/0/0
[PE_1-GigabitEthernet2/0/0] ip address 10.3.1.1 24
[PE_1-GigabitEthernet2/0/0] quit
```

P_1 上的配置。

```
<Huawei> system-view
[Huawei] sysname P_1
[P_1] interface loopback 1
[P_1-LoopBack1] ip address 10.10.1.2 32
[P_1-LoopBack1] quit
[P_1] interface gigabitethernet 1/0/0
[P_1-GigabitEthernet1/0/0] ip address 10.1.1.2 24
[P_1-GigabitEthernet1/0/0] quit
[P_1] interface gigabitethernet 2/0/0
[P_1-GigabitEthernet2/0/0] ip address 10.2.1.1 24
[P_1-GigabitEthernet2/0/0] quit
```

P_2 上的配置。

```
<Huawei> system-view
[Huawei] sysname P_2
[P_2] interface loopback 1
[P_2-LoopBack1] ip address 10.10.1.3 32
[P_2-LoopBack1] quit
[P_2] interface gigabitethernet 1/0/0
[P_2-GigabitEthernet1/0/0] ip address 10.3.1.2 24
[P_2-GigabitEthernet1/0/0] quit
[P_2] interface gigabitethernet 2/0/0
```

```
[P_2-GigabitEthernet2/0/0] ip address 10.4.1.1 24
[P_2-GigabitEthernet2/0/0] quit
```

PE_2 上的配置。

```
<Huawei> system-view
[Huawei] sysname PE_2
[PE_2] interface loopback 1
[PE_2-LoopBack1] ip address 10.10.1.4 32
[PE_2-LoopBack1] quit
[PE_2] interface gigabitethernet 1/0/0
[PE_2-GigabitEthernet1/0/0] ip address 10.2.1.2 24
[PE_2-GigabitEthernet1/0/0] quit
[PE_2] interface gigabitethernet 2/0/0
[PE_2-GigabitEthernet2/0/0] ip address 10.4.1.2 24
[PE_2-GigabitEthernet2/0/0] quit
```

（2）配置 OSPF 路由，把各接口所在网段路由均加入 OSPF 进程 1，区域 0 中（单区域 OSPF 网络时区域 ID 任意），并通过把 PE_1 和 PE_2 的 GE2/0/0 接口的开销设置为 10，其他接口的开销值保持缺省（缺省情况下 GE 接口的开销值为 1），使得通过 P_2 的链路为备用链路。

【经验提示】之所以要把 PE_1 和 PE_2 的 GE2/0/0 接口的开销值增大，是为了使得经过 P_2 的链路为路由备用链路（开销越大，对应的 OSPF 路由表项优先级越低）。OSPF 路由表项优先级是根据链路开销来计算的，本示例两条链路全是 GE 以太网端口，如果全部按缺省开销，这两条链路是等价开销，即等价 OSPF 路由。而链路开销只计算出接口的开销（不计算入接口开销），所以需要分别将 PE_1 和 PE_2 的 GE2/0/0 接口的开销值增大，以使通过 P_2 备用链路从 PE_1 到达 PE_2 和从 PE_2 到达 PE_1 的 OSPF 路由开销都大于通过 P_1 主用链路两个方向的 OSPF 路由开销，最终使其发挥备用路由的作用。但要注意，如果仅提高 PE_1 或 PE_2 中一端的 GE2/0/0 接口的开销值，则仅对一个方向的 OSPF 路由开销起作用。

PE_1 上的配置。

```
[PE_1] ospf 1
[PE_1-ospf-1] area 0
[PE_1-ospf-2-area-0.0.0.0] network 10.10.1.1 0.0.0.0
[PE_1-ospf-2-area-0.0.0.0] network 10.1.1.0 0.0.0.255
[PE_1-ospf-2-area-0.0.0.0] network 10.3.1.0 0.0.0.255
[PE_1-ospf-2-area-0.0.0.0] quit
[PE_1-ospf-1] quit
[PE_1] interface gigabitethernet 2/0/0
[PE_1-GigabitEthernet2/0/0] ospf cost 10
[PE_1-GigabitEthernet2/0/0] quit
```

P_1 上的配置。

```
[P_1] ospf 1
[P_1-ospf-1] area 0
[P_1-ospf-1-area-0.0.0.0] network 10.10.1.2 0.0.0.0
[P_1-ospf-1-area-0.0.0.0] network 10.1.1.0 0.0.0.255
[P_1-ospf-1-area-0.0.0.0] network 10.2.1.0 0.0.0.255
[P_1-ospf-1-area-0.0.0.0] quit
[P_1-ospf-1] quit
```

P_2 上的配置。

```
[P_2] ospf 1
[P_2-ospf-1] area 0
```

```
[P_2-ospf-2-area-0.0.0.0] network 10.10.1.3 0.0.0.0
[P_2-ospf-2-area-0.0.0.0] network 10.3.1.0 0.0.0.255
[P_2-ospf-2-area-0.0.0.0] network 10.4.1.0 0.0.0.255
[P_2-ospf-2-area-0.0.0.0] quit
[P_2-ospf-1] quit
```

PE_2 上的配置。

```
[PE_2] ospf 1
[PE_2-ospf-1] area 0
[PE_2-ospf-2-area-0.0.0.0] network 10.10.1.4 0.0.0.0
[PE_2-ospf-2-area-0.0.0.0] network 10.2.1.0 0.0.0.255
[PE_2-ospf-2-area-0.0.0.0] network 10.4.1.0 0.0.0.255
[PE_2-ospf-2-area-0.0.0.0] quit
[PE_2-ospf-1] quit
[PE_2] interface gigabitethernet 2/0/0
[PE_2-GigabitEthernet2/0/0] ospf cost 10
```

上述配置完成后，在各节点上执行 **display ospf routing** 命令，可以看到相互之间都学到了到达彼此的路由，且从 PE_1 到 PE_2 的 Loopback1 接口的路由出接口为 GE1/0/0（即采用主链路），下一跳为 P_1 的 GE1/0/0 接口 IP 地址 10.1.1.2；从 PE_2 到 PE_1 的 Loopback1 接口的路由出接口为 GE1/0/0（也即采用主链路），下一跳为 P_1 的 GE2/0/0 接口 IP 地址 10.2.1.1（参见输出信息中的粗体字部分）。

```
<PE_1>display ospf routing

    OSPF Process 1 with Router ID 10.1.1.1
            Routing Tables

 Routing for Network
 Destination      Cost  Type      NextHop       AdvRouter     Area
 10.1.1.0/24      1     Transit   10.1.1.1      10.1.1.1      0.0.0.0
 10.3.1.0/24      10    Transit   10.3.1.1      10.1.1.1      0.0.0.0
 10.10.1.1/32     0     Stub      10.10.1.1     10.1.1.1      0.0.0.0
 10.2.1.0/24      2     Transit   10.1.1.2      10.2.1.2      0.0.0.0
 10.4.1.0/24      11    Transit   10.3.1.2      10.3.1.2      0.0.0.0
 10.10.1.2/32     1     Stub      10.1.1.2      10.1.1.2      0.0.0.0
 10.10.1.3/32     10    Stub      10.3.1.2      10.3.1.2      0.0.0.0
 10.10.1.4/32     2     Stub      10.1.1.2      10.2.1.2      0.0.0.0

 Total Nets: 8
 Intra Area: 8   Inter Area: 0   ASE: 0   NSSA: 0

<PE_1>

<PE_2>display ospf routing

    OSPF Process 1 with Router ID 10.2.1.2
            Routing Tables

 Routing for Network
 Destination      Cost  Type      NextHop       AdvRouter     Area
 10.2.1.0/24      1     Transit   10.2.1.2      10.2.1.2      0.0.0.0
 10.4.1.0/24      10    Transit   10.4.1.2      10.2.1.2      0.0.0.0
 10.10.1.4/32     0     Stub      10.10.1.4     10.2.1.2      0.0.0.0
 10.1.1.0/24      2     Transit   10.2.1.1      10.1.1.2      0.0.0.0
```

```
10.3.1.0/24      12    Transit   10.2.1.1      10.3.1.2      0.0.0.0
10.10.1.1/32     2     Stub      10.2.1.1      10.1.1.1      0.0.0.0
10.10.1.2/32     1     Stub      10.2.1.1      10.1.1.2      0.0.0.0
10.10.1.3/32     10    Stub      10.4.1.1      10.3.1.2      0.0.0.0

Total Nets: 8
Intra Area: 8   Inter Area: 0   ASE: 0   NSSA: 0

<PE_2>
```

（3）配置 MPLS 基本功能。LSR ID 以各自的 Loopback1 接口的 IP 地址进行配置，在全局以及 LSP 路径上各 LSR 间互联的接口上使能 MPLS 功能。

PE_1 上的配置。

```
[PE_1] mpls lsr-id 10.10.1.1
[PE_1] mpls
[PE_1-mpls] quit
[PE_1] interface gigabitethernet 1/0/0
[PE_1-GigabitEthernet1/0/0] mpls
[PE_1-GigabitEthernet1/0/0] quit
[PE_1] interface gigabitethernet 2/0/0
[PE_1-GigabitEthernet2/0/0] mpls
[PE_1-GigabitEthernet2/0/0] quit
```

P_1 上的配置。

```
[P_1] mpls lsr-id 10.10.1.2
[P_1] mpls
[P_1-mpls] quit
[P_1] interface gigabitethernet 1/0/0
[P_1-GigabitEthernet1/0/0] mpls
[P_1-GigabitEthernet1/0/0] quit
[P_1] interface gigabitethernet 2/0/0
[P_1-GigabitEthernet2/0/0] mpls
[P_1-GigabitEthernet2/0/0] quit
```

P_2 上的配置。

```
[P_2] mpls lsr-id 10.10.1.3
[P_2] mpls
[P_2-mpls] quit
[P_2] interface gigabitethernet 1/0/0
[P_2-GigabitEthernet1/0/0] mpls
[P_2-GigabitEthernet1/0/0] quit
[P_2] interface gigabitethernet 2/0/0
[P_2-GigabitEthernet2/0/0] mpls
[P_2-GigabitEthernet2/0/0] quit
```

PE_2 上的配置。

```
[PE_2] mpls lsr-id 10.10.1.4
[PE_2] mpls
[PE_2-mpls] quit
[PE_2] interface gigabitethernet 1/0/0
[PE_2-GigabitEthernet1/0/0] mpls
[PE_2-GigabitEthernet1/0/0] quit
[PE_2] interface gigabitethernet 2/0/0
[PE_2-GigabitEthernet2/0/0] mpls
[PE_2-GigabitEthernet2/0/0] quit
```

（4）创建静态 LSP，主、备两条链路均有两条互为反方向的静态 LSP。假设主用链

路中对应的两个方向的LSP名称分别为LSP1、LSP2，备用链路中对应的两个方向的LSP名称分别为LSP3、LSP4。

■ 配置正向静态LSP1，将PE_1作为Ingress，P_1作为Transit，PE_2作为Egress。

Ingress PE_1上的配置。

配置目的IP地址（PE_2的Loopback1接口IP地址）、下一跳（P_1的GE1/0/0接口IP地址）和出标签（假设为20）。

```
[PE_1] static-lsp ingress LSP1 destination 10.10.1.4 32 nexthop 10.1.1.2 out-label 20
```

Transit P_1上的配置。

配置入接口（P_1的GE1/0/0接口）、入标签（20，要与PE_1的出标签一致）、下一跳（PE_2的GE1/0/0接口IP地址）和出标签（假设为40）。

```
[P_1] static-lsp transit LSP1 incoming-interface gigabitethernet 1/0/0 in-label 20 nexthop 10.2.1.2 out-label 40
```

Egress PE_2上的配置。

配置入接口（PE_2的GE1/0/0）和入标签（40，要与P_1的出标签一致）。

```
[PE_2] static-lsp egress LSP1 incoming-interface gigabitethernet 1/0/0 in-label 40
```

■ 配置正向静态LSP2，将PE_1作为Ingress，P_2作为Transit，PE_2作为Egress。

Ingress PE_1上的配置。

配置目的IP地址（PE_2的Loopback1接口IP地址）、下一跳（P_2的GE1/0/0接口IP地址）和出标签（假设为30）。

```
[PE_1] static-lsp ingress LSP2 destination 10.10.1.4 32 nexthop 10.3.1.2 out-label 30
```

Transit P_2上的配置。

配置入接口（P_2的GE1/0/0接口）、入标签（30，要与PE_1的出标签一致）、下一跳（PE_2的GE2/0/0接口IP地址）和出标签（假设为60）。

```
[P_2] static-lsp transit LSP2 incoming-interface gigabitethernet 1/0/0 in-label 30 nexthop 10.4.1.2 out-label 60
```

Egress PE_2上的配置。

配置入接口（PE_2的GE2/0/0）和入标签（60，要与P_2的出标签一致）。

```
[PE_2] static-lsp egress LSP2 incoming-interface gigabitethernet 2/0/0 in-label 60
```

■ 配置反向静态LSP3，将PE_2作为Ingress，P_1作为Transit，PE_1作为Egress。

Ingress PE_2上的配置。

配置目的IP地址（PE_1的Loopback1接口IP地址）、下一跳（P_1的GE2/0/0接口IP地址）和出标签（假设为70）。

```
[PE_2] static-lsp ingress LSP3 destination 10.10.1.1 32 nexthop 10.2.1.1 out-label 70
```

Transit P_1上的配置。

配置入接口（P_1的GE2/0/0接口）、入标签（70，要与PE_2的出标签一致）、下一跳（PE_1的GE1/0/0接口IP地址）和出标签（假设为80）。

```
[P_1] static-lsp transit LSP3 incoming-interface gigabitethernet 2/0/0 in-label 70 nexthop 10.1.1.1 out-label 80
```

Egress PE_1上的配置。

配置入接口（PE_1的GE1/0/0）和入标签（80，要与P_1的出标签一致）。

```
[PE_1] static-lsp egress LSP3 incoming-interface gigabitethernet 1/0/0 in-label 80
```

■ 配置反向静态LSP4，将PE_2作为Ingress，P_2作为Transit，PE_1作为Egress。

Ingress PE_2上的配置。

配置目的IP地址（PE_1的Loopback1接口IP地址）、下一跳（P_2的GE2/0/0接口

IP 地址）和出标签（假设为 90）。

```
[PE_2] static-lsp ingress LSP4 destination 10.10.1.1 32 nexthop 10.4.1.1 out-label 90
```

Transit P_2 上的配置。

配置入接口（P_2 的 GE2/0/0 接口）、入标签（90，要与 PE_2 的出标签一致）、下一跳（PE_1 的 GE2/0/0 接口 IP 地址）和出标签（假设为 100）。

```
[P_2] static-lsp transit LSP4 incoming-interface gigabitethernet 2/0/0 in-label 90 nexthop 10.3.1.1 out-label 100
```

Egress PE_1 上的配置。

配置入接口（PE_1 的 GE2/0/0）和入标签（100，要与 P_2 的出标签一致）。

```
[PE_1] static-lsp egress LSP4 incoming-interface gigabitethernet 2/0/0 in-label 100
```

（5）配置静态 LSP BFD 会话。在此仅以正向静态 LSP1 的配置为例进行介绍，其他 LSP 的 BFD 会话配置方法类似。

在 Ingress PE_1 上配置 BFD 会话，绑定静态 LSP1，假设其中的本地标识符是 1，远端标识符是 2，发送报文的最小时间间隔是 100 毫秒，接收报文的最小时间间隔是 100 毫秒，并且指定反向通道也采用 LSP，能够修改端口状态表。

```
[PE_1] bfd  #---使能 BFD 会话功能
[PE_1-bfd] quit
[PE_1] bfd pe1tope2 bind static-lsp LSP1  #---创建 BFD 会话，并指定绑定静态 LSP1
[PE_1-bfd-lsp-session-pe1tope2] discriminator local 1  #---配置本地标识符为 1
[PE_1-bfd-lsp-session-pe1tope2] discriminator remote 2 #---配置远程标识符为 1
[PE_1-bfd-lsp-session-pe1tope2] min-tx-interval 100  #---配置发送 BFD 会话报文的最小时间间隔为 100 毫秒
[PE_1-bfd-lsp-session-pe1tope2] min-rx-interval 100  #---配置接收 BFD 会话报文的最小时间间隔为 100 毫秒
[PE_1-bfd-lsp-session-pe1tope2] process-pst  #---使反向 LSP 通道在当前 BFD 会话 Down 时也进行主备切换
[PE_1-bfd-lsp-session-pe1tope2] commit  #---提交以上配置，使配置生交效
[PE_1-bfd-lsp-session-pe1tope2] quit
```

在 Egress PE_2 上配置 BFD 会话，通过 IP 链路向 Ingress PE_1 通告静态 LSP1 故障。本地中 PE_1 和 PE_2 不是直接连接的，不配置 **process-pst** 命令，使反射通道为 IP 路由。

```
[PE_2] bfd
[PE_2-bfd] quit
[PE_2] bfd pe2tope1 bind peer-ip 10.10.1.1
[PE_2-bfd-session-pe2tope1] discriminator local 2
[PE_2-bfd-session-pe2tope1] discriminator remote 1
[PE_2-bfd-session-pe2tope1] min-tx-interval 100
[PE_2-bfd-session-pe2tope1] min-rx-interval 100
[PE_2-bfd-session-pe2tope1] commit
[PE_2-bfd-session-pe2tope1] quit
```

3. 实验结果验证

以上配置完成后，下面来依次验证静态 LSP 和 BFD 检测配置效果。

（1）验证静态 LSP 配置。

在各节点上用 **display mpls static-lsp verbose** 命令可查看本地设备上所配置的静态 LSP 的详细信息。以下是分别在 PE_1 和 PE_2 上执行 **display mpls static-lsp verbose** 命令的输出，显示了所创建的 4 条静态 LSP 的主要参数信息，其中：LSP1 和 LSP2 以 PE_1 作为 Ingress，PE_2 作为 Egress；LSP3 和 LSP4 以 PE_2 作为 Ingress，PE_1 作为 Egress。

```
[PE_1] display mpls static-lsp verbose
No              : 1
LSP-Name        : LSP1
```

```
LSR-Type          : Ingress
FEC               : 10.10.1.4/32
In-Label          : NULL
Out-Label         : 20
In-Interface      : -
Out-Interface     : GigabitEthernet1/0/0
NextHop           : 10.1.1.2
Static-Lsp Type: Normal
Lsp Status        : Up

No                : 2
LSP-Name          : LSP2
LSR-Type          : Ingress
FEC               : 10.10.1.4/32
In-Label          : NULL
Out-Label         : 30
In-Interface      : -
Out-Interface     : GigabitEthernet2/0/0
NextHop           : 10.3.1.2
Static-Lsp Type: Normal
Lsp Status        : Down

No                : 3
LSP-Name          : LSP3
LSR-Type          : Egress
FEC               : 10.10.1.1/32
In-Label          : 80
Out-Label         : NULL
In-Interface      :  GigabitEthernet1/0/0
Out-Interface     : -
NextHop           : -
Static-Lsp Type: Normal
Lsp Status        : Up

No                : 4
LSP-Name          : LSP4
LSR-Type          : Egress
FEC               : 10.10.1.1/32
In-Label          : 100
Out-Label         : NULL
In-Interface      : GigabitEthernet2/0/0
Out-Interface     : -
NextHop           : -
Static-Lsp Type: Normal
Lsp Status        : Up

<PE_1>

<PE_2>display mpls static-lsp verbose
No                : 1
LSP-Name          : LSP1
LSR-Type          : Egress
FEC               : -/-
In-Label          :  40
```

```
Out-Label         :  NULL
In-Interface      :  GigabitEthernet1/0/0
Out-Interface     :  -
NextHop           :  -
Static-Lsp Type:     Normal
Lsp Status        :  Up

No                : 2
LSP-Name          : LSP2
LSR-Type          : Egress
FEC               : -/-
In-Label          :  60
Out-Label         :  NULL
In-Interface      :  GigabitEthernet2/0/0
Out-Interface     :  -
NextHop           :  -
Static-Lsp Type:     Normal
Lsp Status        :  Up

No                : 3
LSP-Name          : LSP3
LSR-Type          : Ingress
FEC               : 10.10.1.1/32
In-Label          :  NULL
Out-Label         :  70
In-Interface      :  -
Out-Interface     :  GigabitEthernet1/0/0
NextHop           : 10.2.1.1
Static-Lsp Type:     Normal
Lsp Status        :  Up

No                : 4
LSP-Name          : LSP4
LSR-Type          : Ingress
FEC               : 10.10.1.1/32
In-Label          :  NULL
Out-Label         :  90
In-Interface      :  -
Out-Interface     :  GigabitEthernet2/0/0
NextHop           : 10.4.1.1
Static-Lsp Type:     Normal
Lsp Status        :  Down

<PE_2>
```

从以上 PE_1 和 PE_2 的输出可以看出，只有 PE_1 的 LSP2 和 PE_2 的 LSP4 的状态为 Down，没有出现预期中 PE_2 的 LSP2 和 PE_1 的 LSP4 的状态为 Down。这与不同节点的静态 LSP 所需配置的参数有关。

在静态 LSP 配置中，只有 Ingress 上才需要配置目的 IP 地址（与 FEC 相关），其他设备均不需要。这样在 Ingress 上会查找对应 FEC 的 IP 路由表项，以验证所配置的下一跳是否可达。而在本示例中，因为已把备用链路上 PE_1 和 PE_2 的 GE2/0/0 接口的 OSPF 路由开销调高，使得在 PE_1 的 IP 路由表中没有经过 P_2 到达 PE_2 的路由表项，在 PE_2 的路由表中没有经过 P_2 到达 PE_1 的路由表项，所以 PE_1 上的 LSP2 和 PE_4 上的 LSP4

配置的下一跳即变成不可达，最终导致这两条静态LSP建立不成功。

在Transit和Egress上，因为不需配置目的IP地址，不会查看与FEC相关的路由表项来验证下一跳是否可达，只会验证本地MPLS接口状态（LSP只有本地意义），所以当其接口没有Down时，其建立的静态LSP就一定是Up状态。

以下是分别在P_1、P_2上执行 **display mpls static-lsp verbose** 命令的输出，它们都作为Transit，且LSP都是Up状态。

```
<P_1>display mpls static-lsp verbose
No              : 1
LSP-Name        : LSP1
LSR-Type        : Transit
FEC             : -/-
In-Label        :  20
Out-Label       :  40
In-Interface    :  GigabitEthernet0/0/0
Out-Interface   :  GigabitEthernet0/0/1
NextHop         : 10.2.1.2
Static-Lsp Type:  Normal
Lsp Status      :  Up

No              : 2
LSP-Name        : LSP3
LSR-Type        : Transit
FEC             : -/-
In-Label        :  70
Out-Label       :  80
In-Interface    :  GigabitEthernet0/0/1
Out-Interface   :  GigabitEthernet0/0/0
NextHop         : 10.1.1.1
Static-Lsp Type:  Normal
Lsp Status      :  Up

<P_1>

<P_2>display mpls static-lsp verbose
No              : 1
LSP-Name        : LSP2
LSR-Type        : Transit
FEC             : -/-
In-Label        :  30
Out-Label       :  60
In-Interface    :  GigabitEthernet0/0/0
Out-Interface   :  GigabitEthernet0/0/1
NextHop         : 10.4.1.2
Static-Lsp Type:  Normal
Lsp Status      :  Up

No              : 2
LSP-Name        : LSP4
LSR-Type        : Transit
FEC             : -/-
In-Label        :  90
Out-Label       :  100
```

```
In-Interface    :   GigabitEthernet0/0/1
Out-Interface   :   GigabitEthernet0/0/0
NextHop         :   10.3.1.1
Static-Lsp Type:    Normal
Lsp Status      :   Up

<P_2>
```

（2）验证 LSP 的连通性及数据转发路径。

在 PE_1 上执行 **ping lsp ip** 10.10.1.4 32 命令，可以 Ping 通，当然，在 PE_2 上执行 **ping lsp ip** 10.10.1.1 32 命令，也可以 Ping 通。

在 PE_1 上执行 **tracert lsp ip** 10.10.1.4 32 命令，或在 PE_2 上执行 **tracert lsp ip** 10.10.1.1 32 命令，可以看到 PE_1 与 PE_2 的通信路径采用的是主用链路。具体如下。

```
<PE_1>tracert lsp ip 10.10.1.4 32
  LSP Trace Route FEC: IPV4 PREFIX 10.10.1.4/32 , press CTRL_C to break.
  TTL    Replier             Time      Type        Downstream
  0                                    Ingress     10.1.1.2/[20 ]
  1      10.1.1.2            40 ms     Transit     10.2.1.2/[40 ]
  2      10.10.1.4           50 ms     Egress
<PE_1>

<PE_2>tracert lsp ip 10.10.1.1 32
  LSP Trace Route FEC: IPV4 PREFIX 10.10.1.1/32 , press CTRL_C to break.
  TTL    Replier             Time      Type        Downstream
  0                                    Ingress     10.2.1.1/[70 ]
  1      10.2.1.1            20 ms     Transit     10.1.1.1/[80 ]
  2      10.10.1.1           50 ms     Egress
<PE_2>
```

（3）验证 BFD 会话状态。

在 PE_1、PE_2 上分别执行 **display bfd session all** 命令可查看前面创建的针对 LSP1 的 BFD 会话建立状态，可以看到 PE_1 和 PE_2 上的 BFD 会话已经 Up。

```
[PE_1] display bfd session all
--------------------------------------------------------------------------------
Local Remote     PeerIpAddr      State     Type          InterfaceName
--------------------------------------------------------------------------------
1     2          10.10.1.4       Up        S_STA_LSP     GigabitEthernet1/0/0
--------------------------------------------------------------------------------
      Total Up/DOWN Session Number : 1/0

[PE_2] display bfd session all
--------------------------------------------------------------------------------
Local Remote     PeerIpAddr      State     Type          InterfaceName
--------------------------------------------------------------------------------
2     1          10.10.1.1       Up        S_IP_PEER          -
--------------------------------------------------------------------------------
      Total Up/DOWN Session Number : 1/0
```

对 P_1 的 GE2/0/0 接口进行 shutdown 操作，模拟静态 LSP 故障。

```
[P_1] interface gigabitethernet 2/0/0
[P_1-GigabitEthernet2/0/0] shutdown
```

在 PE_1 和 PE_2 上分别执行 **display bfd session all** 命令，查看 BFD 的会话状态，发现前面创建的针对 LSP1 的 BFD 会话均呈现 Down 状态，因为 P_1 的接口关闭了，链

路不通。

```
[PE_1] display bfd session all
--------------------------------------------------------------------------------
Local Remote      PeerIpAddr        State       Type           InterfaceName
--------------------------------------------------------------------------------
1       2           10.10.1.4          Down        S_STA_LSP            -
--------------------------------------------------------------------------------
      Total Up/DOWN Session Number : 0/1

[PE_2] display bfd session all
--------------------------------------------------------------------------------
Local Remote      PeerIpAddr        State       Type           InterfaceName
--------------------------------------------------------------------------------
2       1           10.10.1.1          Down        S_IP_PEER            -
--------------------------------------------------------------------------------
      Total Up/DOWN Session Number : 0/1
```

如果此时再在 PE_1 或 PE_2 上执行 **display mpls static-lsp verbose** 命令，会发现原来呈 Up 状态的两条主用链路上的静态 LSP 状态变为 Down 状态；相反，原来呈 Down 状态的两条备用链路上的静态 LSP 状态变为 Up 状态。这同样是因为已没有经过 P_1 的路由表项了。

```
<PE_1>display mpls static-lsp verbose
No                  : 1
LSP-Name            : LSP1
LSR-Type            : Ingress
FEC                 : 10.10.1.4/32
In-Label        :   NULL
Out-Label       :   20
In-Interface    :   -
Out-Interface   :   GigabitEthernet1/0/0
NextHop             : 10.1.1.2
Static-Lsp Type:    Normal
Lsp Status      :   Down

No                  : 2
LSP-Name            : LSP2
LSR-Type            : Ingress
FEC                 : 10.10.1.4/32
In-Label        :   NULL
Out-Label       :   30
In-Interface    :   -
Out-Interface   :   GigabitEthernet2/0/0
NextHop             : 10.3.1.2
Static-Lsp Type:    Normal
Lsp Status      :   Up

No                  : 3
LSP-Name            : LSP3
LSR-Type            : Egress
FEC                 : -/-
In-Label        :   80
Out-Label       :   NULL
In-Interface    :   GigabitEthernet1/0/0
Out-Interface   :   -
```

```
NextHop          : -
Static-Lsp Type:   Normal
Lsp Status     :   Up

No               : 4
LSP-Name         : LSP4
LSR-Type         : Egress
FEC              : -/-
In-Label       :   100
Out-Label      :   NULL
In-Interface   :   GigabitEthernet2/0/0
Out-Interface  :   -
NextHop          : -
Static-Lsp Type:   Normal
Lsp Status     :   Up

<PE_1>

<PE_2>display mpls static-lsp verbose
No               : 1
LSP-Name         : LSP1
LSR-Type         : Egress
FEC              : -/-
In-Label       :   40
Out-Label      :   NULL
In-Interface   :   GigabitEthernet1/0/0
Out-Interface  :   -
NextHop          : -
Static-Lsp Type:   Normal
Lsp Status     :   Down

No               : 2
LSP-Name         : LSP2
LSR-Type         : Egress
FEC              : -/-
In-Label       :   60
Out-Label      :   NULL
In-Interface   :   GigabitEthernet2/0/0
Out-Interface  :   -
NextHop          : -
Static-Lsp Type:   Normal
Lsp Status     :   Up

No               : 3
LSP-Name         : LSP3
LSR-Type         : Ingress
FEC              : 10.10.1.1/32
In-Label       :   NULL
Out-Label      :   70
In-Interface   :   -
Out-Interface  :   GigabitEthernet1/0/0
NextHop          : 10.2.1.1
Static-Lsp Type:   Normal
Lsp Status     :   Down
```

```
No             : 4
LSP-Name       : LSP4
LSR-Type       : Ingress
FEC            : 10.10.1.1/32
In-Label       :  NULL
Out-Label      :  90
In-Interface   :  -
Out-Interface  :  GigabitEthernet2/0/0
NextHop        : 10.4.1.1
Static-Lsp Type:  Normal
Lsp Status     :  Up

<PE_2>
```

在 P_1 上执行 **display mpls static-lsp verbose** 命令会发现两条 LSP 均为 Down 状态，那是因为它的 GE2/0/0 接口被关闭，呈 Down 状态了，而在在 P_2 上执行 **display mpls static-lsp verbose** 命令会发现两条 LSP 均为 Up 状态。

```
<P_1>display mpls static-lsp verbose
No             : 1
LSP-Name       : LSP1
LSR-Type       : Transit
FEC            : -/-
In-Label       :  20
Out-Label      :  40
In-Interface   :  GigabitEthernet1/0/0
Out-Interface  :  GigabitEthernet2/0/0
NextHop        : 10.2.1.2
Static-Lsp Type:  Normal
Lsp Status     :  Down

No             : 2
LSP-Name       : LSP3
LSR-Type       : Transit
FEC            : -/-
In-Label       :  70
Out-Label      :  80
In-Interface   :  GigabitEthernet2/0/0
Out-Interface  :  GigabitEthernet1/0/0
NextHop        : 10.1.1.1
Static-Lsp Type:  Normal
Lsp Status     :  Down

<P_1>

<P_2>display mpls static-lsp verbose
No             : 1
LSP-Name       : LSP2
LSR-Type       : Transit
FEC            : -/-
In-Label       :  30
Out-Label      :  60
In-Interface   :  GigabitEthernet1/0/0
Out-Interface  :  GigabitEthernet2/0/0
NextHop        : 10.4.1.2
Static-Lsp Type:  Normal
```

```
Lsp Status        :   Up

No                : 2
LSP-Name          : LSP4
LSR-Type          : Transit
FEC               : -/-
In-Label          :   90
Out-Label         :   100
In-Interface      :   GigabitEthernet2/0/0
Out-Interface     :   GigabitEthernet1/0/0
NextHop           : 10.3.1.1
Static-Lsp Type:      Normal
Lsp Status        :   Up
<P_2>
```

如果查看从 PE_1 到达 PE_2 的 Loobpack1 接口 OSPF 路由，会发现采用的下一跳是 P_2 的 GE1/0/0 接口。查看从 PE_2 到达 PE_1 的 Loobpack1 接口 OSPF 路由，会发现采用的下一跳是 P_2 的 GE2/0/0 接口（参见输出信息中的粗体字部分）。

```
<PE_1>display ospf routing

   OSPF Process 1 with Router ID 10.1.1.1
           Routing Tables

 Routing for Network
 Destination        Cost  Type       NextHop        AdvRouter       Area
 10.1.1.0/24        1     Transit    10.1.1.1       10.1.1.1        0.0.0.0
 10.3.1.0/24        1     Transit    10.3.1.1       10.1.1.1        0.0.0.0
 10.10.1.1/32       0     Stub       10.10.1.1      10.1.1.1        0.0.0.0
 10.4.1.0/24        2     Transit    10.3.1.2       10.3.1.2        0.0.0.0
 10.10.1.2/32       1     Stub       10.1.1.2       10.1.1.2        0.0.0.0
 10.10.1.3/32       1     Stub       10.3.1.2       10.3.1.2        0.0.0.0
 10.10.1.4/32       2     Stub       10.3.1.2       10.2.1.2        0.0.0.0

 Total Nets: 7
 Intra Area: 7   Inter Area: 0   ASE: 0   NSSA: 0

<PE_1>

<PE_2>display ospf routing

   OSPF Process 1 with Router ID 10.2.1.2
           Routing Tables

 Routing for Network
 Destination        Cost  Type       NextHop        AdvRouter       Area
 10.2.1.0/24        1     Transit    10.2.1.2       10.2.1.2        0.0.0.0
 10.4.1.0/24        10    Transit    10.4.1.2       10.2.1.2        0.0.0.0
 10.10.1.4/32       0     Stub       10.10.1.4      10.2.1.2        0.0.0.0
 10.1.1.0/24        2     Transit    10.2.1.1       10.1.1.2        0.0.0.0
 10.3.1.0/24        12    Transit    10.2.1.1       10.3.1.2        0.0.0.0
 10.10.1.1/32       2     Stub       10.2.1.1       10.1.1.1        0.0.0.0
 10.10.1.2/32       1     Stub       10.2.1.1       10.1.1.2        0.0.0.0
 10.10.1.3/32       10    Stub       10.4.1.1       10.3.1.2        0.0.0.0
```

```
Total Nets: 8
Intra Area: 8   Inter Area: 0   ASE: 0   NSSA: 0

<PE_2>
```

2.2　静态 LSP 建立不成功故障排除

虽然静态 LSP 的配置很简单，但也可能因为一些细节配置问题，造成最终的 LSP 建立不成功，下面具体介绍这种故障的排除方法。

（1）首先利用 **display mpls static-lsp** 命令查看 LSP 的状态，以下是一个输出示例。

```
<PE_1>dis mpls static-lsp
TOTAL          :    4        STATIC LSP(S)
Up             :    2        STATIC LSP(S)
DOWN           :    2        STATIC LSP(S)
Name           FEC              I/O Label     I/O If          Status
LSP1           10.10.1.4/32     NULL/20       -/GE0/0/0       Down
LSP2           10.10.1.4/32     NULL/30       -/GE0/0/1       Up
LSP3           -/-              80/NULL       GE0/0/0/-       Down
LSP4           -/-              100/NULL      GE0/0/1/-       Up
```

如果发现某条 LSP 的状态（Status）为 Down（如以上的 LSP1 和 LSP3），证明其没有最终建立成功。

（2）因为静态 LSP 的状态只与本地配置有关，所以首先要检查这些呈 Down 状态的 LSP 在本地的配置，包括用于标识 LSR ID 的 Loopback 接口 IP 地址配置。还要看设备间相连的接口上是否启用了 MPLS 功能，当然首先要在全局使能 MPLS 功能。

（3）还可在对应的 LSP 路径各设备上分别执行 **display mpls static-lsp** *lsp-name* **verbose** 命令，查看对应静态 LSP 的详细配置。如在上一步发现 PE_1 的 LSP1 为 Down 状态，就要在 PE_1 上执行 **display mpls static-lsp** LSP1 **verbose**（注意，LSP 名称区分大小写），显示这条 LSP 在本地设备上的详细配置，具体如下。

```
<PE_1>dis mpls static-lsp LSP1 verbose
No                 : 1
LSP-Name           : LSP1
LSR-Type           : Ingress
FEC                : 10.10.1.4/32
In-Label       :   NULL
Out-Label      :   20
In-Interface   :   -
Out-Interface  :   GigabitEthernet0/0/0
NextHop            : 10.1.1.2
Static-Lsp Type:   Normal
Lsp Status     :   Down
```

从输出信息中可以看出本条 LSP 在本地设备的详细配置（不同节点类型所需配置的参数不完全一样）。当然，要验证 LSP 的配置是否正确，必须在路径的所有设备上执行以上命令进行配置查看。

在这里要特别注意，**一定要确保 LSP 路径的上游节点的出标签要与下游节点的入标签保持一致**。另外，静态 LSP 的标签取值范围在 16~1023 之间，当然在 Egress 上可

以根据需要使用诸如 0、2、3 这样的特殊标签作为入标签（倒数第二跳的出标签值要与其一致）。

一般情况下，Ingress 上只有出标签，Egress 上只有入标签，而 Transit 上必须同时有入标签和出标签。当然，在 Ingress 和 Transit 上配置的下一跳也必须正确，否则可能造成路径不通。

（4）如果各设备上的静态 LSP 标签配置没有问题，但 LSP 仍为 Down 状态，就要考虑路径中的某个出接口是否被路由协议管理 Down 掉了（成为备份出接口），如在 1.2.6 小节介绍的主、备链路中，备份链路就被 OSPF 协议管理 Down 掉了。还要考虑入接口或出接口是否被关闭了，因为这样即使静态 LSP 配置完全正确，也会显示 Down 状态。





第3章 MPLS LDP基本功能配置与管理



第 2 章介绍了手工配置方式的静态 LSP 的建立配置与管理方法，本章要介绍采用 LDP 通过动态协商建立 LSP 的配置与管理方法。这两章所介绍的内容都是为了在 MPLS 骨干网中构建将用于数据传输的 LSP 隧道。

LDP 是 MPLS 体系中非常重要的标签发布控制协议。如果把静态 LSP 比作静态路由的话，那么 LDP 就相当于一种动态路由协议，即无需网络维护人员在各节点上手工一条条去配置 LSP，通过 LDP 就可以在各节点上动态建立 LSP，极大地减轻了维护人员的工作量，同时也减少了配置错误的发生。

本章主要围绕 LDP LSP 建立过程中所涉及的一些基本功能的配置与管理方法，以及 LDP LSP 建立典型故障的排除方法进行介绍。LDP 基本又分必选基本功能和可选基本功能，必选基本功能的配置其实就是配置在对等体之间的 LDP 会话（包括本地会话和远端会话），因为只要在对等体间建立好了 LDP 会话，就可以成功建立 LSP。LDP 可选基本功能比较丰富，涉及各种 LDP 会话定时器、标签发布和分配控制方式、DoD 请求自动触发，以及多种标签映射消息过滤（类似于路由信息过滤）、LSP 建立的触发策略等。

3.1 LDP 基础及工作原理

LDP（Label Distribution Protocol，标签分发协议）是 MPLS 的一种最主要的控制协议，负责 FEC 的分类、MPLS 标签的分配以及 LSP 的动态建立和维护等操作。

LDP 规定了标签分发过程中的各种消息以及相关处理过程。通过 LDP 协议，LSR 可以把网络层的路由信息直接映射到数据链路层的 LSP 交换路径上，实现在网络层动态建立 LSP。目前，LDP 广泛地应用在 VPN 服务上，具有组网、配置简单，支持基于路由动态建立 LSP，支持大容量 LSP 等优点。

3.1.1 LDP 基本概念

在利用 LDP 动态建立 LSP 的过程中，涉及以下基本概念。

1. LDP 对等体

LDP 对等体是指相互之间存在直接的 LDP 会话、可直接使用 LDP 来交换标签消息（包括标签请求消息和标签映射消息）的两个 LSR。比如，BGP 路由协议和 IKE（因特网密钥交换）中的对等体，其实它们的基本设计思想是一样的，尽管它们所实现的功能不一样。在 LDP 对等体中，通过它们之间的 LDP 会话可获得下游对等体为某 FEC 分配的 MPLS 入标签，以此作为本端对应 FEC 的出标签。LDP 对等体之间可以是直连的，也可以是非直连的，这点与 IBGP 中的对等体类似（MP-EBGP 中的对等体也可以是非直连的）。

如图 3-1 所示，LSR_1 下面连接了一台二层交换机 SW，然后在这台二层交换机下又连接多个 LSR 路由器，LSR_1 与 LSR_2、LSR_3 和 LSR_4 之间则可以看成是直连的对等体关系；同理，LSR_5 与 LSR_2、LSR_3 和 LSR_4 之间也是直连对等体关系，但 LSR_1 与 LSR_5 之间也可以建立对等体关系，它们之间是非直连对等体关系。

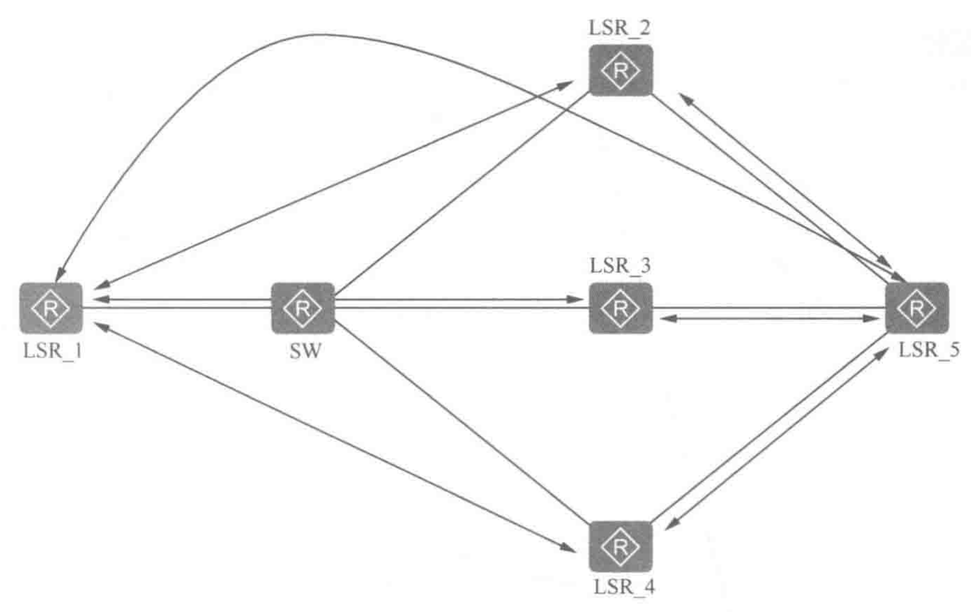


图 3-1 LSR 对等体示例

2. LDP 邻接体

当一台LSR接收到对端发送的Hello消息后，两端之间就建立了LDP邻接体关系（也就是邻居关系），这种LDP邻接体存在两种类型。

■ 本地邻接体（Local Adjacency）：通过组播形式（目的IP地址为224.0.0.2，代表本地子网中的所有路由器）发送Hello消息（称为"链路Hello消息"）发现的邻接体叫作本地邻接体。

当一个源路由器通过一接口以组播方式发送一条Hello消息后，则在这条链路下的所有直连路由器都会收到，然后这些路由器就是这个源路由器的本地邻接体。如图3-1中的LSR_1以组播方式发送一条Hello消息，则LSR_2、LSR_3和LSR_4都可以收到，它们都是LSR_1的本地邻接体。

■ 远端邻接体（Remote Adjacency）：以单播形式发送Hello消息（称之为"目标Hello消息"）发现的邻接体叫作远端邻接体。

远端邻接体通常被认为是非直连的，但也可以是直连的。如图3-1中的LSR_1向LSR_5以单播方式发送一条Hello消息，则LSR_5就是LSR_1的远端邻接体。

LDP通过邻接体来维护对等体的存在，对等体的类型取决于维护它的邻接体的类型。一个对等体可以由多个邻接体来维护，如果同时包括本地邻接体和远端邻接体，则该对等体类型为本远共存对等体。

3. LDP 会话

LDP会话用于在LSR之间交换标签映射、释放会话等消息。只有存在邻接体的两端对等体之间才能建立LDP会话。区分于本地邻接体和远端邻接体，在这两类邻接体之间建立的LDP会话也对应分为两种类型。

■ 本地LDP会话（Local LDP Session）：建立会话的两个LSR之间是直连的本地邻接体关系。

■ 远端LDP会话（Remote LDP Session）：建立会话的两个LSR之间可以是直连的

本地邻接体关系，也可以是非直连的远端邻接体关系。

本地 LDP 会话和远端 LDP 会话可以共存，也就是一个对等体上可以同时创建与直连对等体的会话，以及与非直连对等体的会话。

3.1.2 LDP 会话消息和两个阶段

LDP 协议规定了标签分发过程中的各种消息以及相关的处理过程。通过 LDP，LSR 可以把网络层的路由信息映射到数据链路层的交换路径上，进而建立起 LSP。

1. LDP 会话消息

在 LDP 会话过程中，主要使用以下 4 类消息。

■ 发现（Discovery）消息：用于通告和维护网络中 LSR 的存在，如 Hello 消息。

■ 会话（Session）消息：用于建立、维护和终止 LDP 对等体之间的会话，如 Initialization（初始化）消息、Keepalive（保持活跃）消息。

■ 通告（Advertisement）消息：用于创建、改变和删除 FEC 的标签映射，如标签映射消息。

■ 通知（Notification）消息：用于提供建议性的消息和差错通知。

为保证 LDP 消息的可靠发送，除了 Discovery 消息使用 UDP 传输外，LDP 的 Session 消息、Advertisement 消息和 Notification 消息都使用 TCP 传输。在所使用的传输层端口上，要区分以下几种情况。

■ Hello 消息都使用 UDP 协议传输，源端口和目的端口均为 646（LDP 协议端口号）。

■ 在 LDP 会话、通告和通知消息中，主动方（对等体间 IP 地址大的一方，下节有介绍）发送的消息中的源端口为任意 TCP 端口，目的端口为 TCP 646（LDP 协议端口号），被动方发送的消息中的源端口为 TCP 646 端口，目的端口为任意 TCP 端口。

2. LDP 的两个工作阶段

LDP 工作过程主要分为两个阶段：先在对等体间建立 LDP 会话，然后才能在对等体间建立 LSP。

（1）LDP 会话的建立

在这个过程中，LSR 设备通过发送 Hello 消息来发现邻接体，然后在 LSR 之间建立 LDP 会话。会话建立后，LDP 对等体之间通过周期性地发送 Hello 消息和 Keepalive 消息来保持这个会话。

■ LDP 对等体之间，通过周期性发送 Hello 消息表明自己希望继续维持这种邻接关系。如果 Hello 保持定时器超时仍没有收到对端发来的新的 Hello 消息，则会删除它们之间的邻接关系。邻接关系被删除后，本端 LSR 将向对端发送 Notification 消息，结束它们之间的 LDP 会话。

■ LDP 对等体之间，通过 LDP 会话连接传送的 Keepalive 消息来维持 LDP 会话。如果会话保持定时器（Keepalive 保持定时器）超时仍没有收到对端发来的新的 Keepalive 消息，则本端 LSR 将向对端发送 Notification 消息，关闭它们之间的 TCP 连接，结束 LDP 会话。

有关 LDP 会话的具体建立流程将在 3.1.3 节介绍。

（2）LDP LSP 的建立

LDP 会话建立成功后，LDP 通过发送标签请求和标签映射消息，在 LDP 对等体之

间通告 FEC 和标签的绑定关系，从而建立 LSP。

有关 LDP 动态 LSP 建立的具体流程将在 3.1.5 节介绍。

3.1.3 LDP 会话的建立流程

通过 LDP 发现机制发现 LDP 对等体后，就可在对等体之间建立 LDP 会话。只有建立了 LDP 会话后，才能建立 LDP LSP 来承载业务。

1. LDP 发现机制

LDP 有两种用于 LSR 发现潜在的 LDP 对等体的机制。

（1）基本发现机制：用于发现直连链路上的 LSR。

LDP 基本发现机制是 LSR 通过周期性地**以组播方式**发送 LDP 链路 Hello 消息（LDP Link Hello）进行的，发现的是直连链路上的 LDP 对等体，并与之建立本地 LDP 会话。

LDP 链路 Hello 消息使用 UDP 协议传输，目的 IP 地址是组播地址 224.0.0.2，源/目的端口均为 UDP 646。如果 LSR 在特定接口接收到邻居 LSR 发来的 LDP 链路 Hello 消息，表明该接口存在 LDP 对等体。

（2）扩展发现机制：用于发现非直连链路上的 LSR。

扩展发现机制是 LSR 周期性地**以单播方式**发送 LDP 目标 Hello 消息（LDP Targeted Hello）到指定 IP 地址进行的，发现的是非直连链路上的 LDP 对等体，并与之建立远端 LDP 会话。

LDP 目标 Hello 消息也使用 UDP 协议传输，目的 IP 地址是指定的对端单播 IP 地址，源/目的端口均为 UDP 646。如果 LSR 接收到 LDP 目标 Hello 消息，表明该 LSR 存在 LDP 对等体。

2. LDP 会话的建立过程

在 LSR 之间建立 LDP 会话的过程总体可以划分三个阶段：第一阶段是通过交互 Hello 消息，相互建立 TCP 连接；第二阶段是通过交互 LDP 会话初始化消息（Initiazation Message），协商会话参数；第三阶段是相互交互 Keepalive 消息，建立 LDP 会话。具体流程如图 3-2 所示，各步骤说明如下（对应图中的步骤）。



图 3-2 LDP 会话的建立流程

（1）两个 LSR 之间互相发送 Hello 消息，基于不同发现机制采用不同的发送方式。双方使用 Hello 消息“源 IP 地址”字段中填充的 IP 地址（称之为“传输地址”）进行 LDP 会话建立。

（2）传输地址较大的一方作为主动方，发起建立 TCP 连接。

如图 3-2 所示，LSR_1 的 IP 地址大于 LSR_2 的 IP 地址，故 LSR_1 作为主动方发起建立 TCP 连接，LSR_2 作为被动方等待对方发起连接。

（3）TCP 连接建立成功后，首先由主动方 LSR_1 向被动方 LSR_2 发送初始化消息（源端口任意，目的端口为 TCP 646），协商建立 LDP 会话的相关参数。

初始化消息中包括 LDP 会话的相关参数，如 LDP 协议版本（Session Protocol Version）、会话标签分发方式（Session Label Advertisement Discipline）、会话 Keepalive 保持定时器（Session KeepAlive Time）、会话环路检测（Session Loop Detection）功能是否启用（缺省不启用），最大 PDU（Seesion Max PDU Length）、会话接收方 LSR ID（Session Receiver LSR Identifier）和会话接收方标签空间（Session Recevier Label Space Identifier，缺省为 0）等，如图 3-3 所示。

```
⊟ Initialization Message
    0... .... = U bit: Unknown bit not set
    Message Type: Initialization Message (0x200)
    Message Length: 22
    Message ID: 0x0000001b
  ⊟ Common Session Parameters TLV
      00.. .... = TLV Unknown bits: Known TLV, do not Forward (0x00)
      TLV Type: Common Session Parameters TLV (0x500)
      TLV Length: 14
    ⊟ Parameters
        Session Protocol Version: 1
        Session KeepAlive Time: 45
        0... .... = Session Label Advertisement Discipline: Downstream Unsolicited proposed
        .0.. .... = Session Loop Detection: Loop Detection Disabled
        Session Path Vector Limit: 0
        Session Max PDU Length: 4096
        Session Receiver LSR Identifier: 3.3.3.3 (3.3.3.3)
        Session Receiver Label Space Identifier: 0
```

图 3-3 LDP 初始化消息示例

（4）被动方 LSR_2 收到初始化消息后，如果接受 LSR_1 发来的相关初始化参数，则自己也向主动方发送初始化消息和 Keepalive 消息，如图 3-4 所示。发送的会话初始化消息中所包括的参数与图 3-3 相同。

在 Keepalive 消息中主要包括消息类型（Message Type，此处为 Keep Alive Message，值为十六进制的 201）、消息长度（Message Length，为 4 个字节）和消息 ID（Message ID）。

如果被动方 LSR_2 不能接受相关初始化参数，则发送 Notification 消息终止 LDP 会话的建立。

（5）主动方 LSR_1 收到被动初始化和 KeepAlive 消息后，如果接受 LSR_2 发来的相关初始化参数，则向被动方发送 KeepAlive 消息和地址消息（Address Message），如图 3-5 所示。KeepAlive 消息所包括的内容与前面介绍的一样，地址消息是仅当双方通过初始化消息最终协商采用 DU（下游自主方式，3.1.4 节将介绍）标签发布方式时双方才进行交互，以便向对方通告本地直连的 32 位掩码 IP 地址和启用了 LDP 协议的接口 IP 地址。

```
⊟ Initialization Message
    0... .... = U bit: Unknown bit not set
    Message Type: Initialization Message (0x200)
    Message Length: 22
    Message ID: 0x00000033
  ⊟ Common Session Parameters TLV
      00.. .... = TLV Unknown bits: Known TLV, do not Forward (0x00)
      TLV Type: Common Session Parameters TLV (0x500)
      TLV Length: 14
    ⊟ Parameters
        Session Protocol Version: 1
        Session KeepAlive Time: 45
        0... .... = Session Label Advertisement Discipline: Downstream Unsolicited proposed
        .0.. .... = Session Loop Detection: Loop Detection Disabled
        Session Path Vector Limit: 0
        Session Max PDU Length: 4096
        Session Receiver LSR Identifier: 5.5.5.5 (5.5.5.5)
        Session Receiver Label Space Identifier: 0
⊟ Keep Alive Message
    0... .... = U bit: Unknown bit not set
    Message Type: Keep Alive Message (0x201)
    Message Length: 4
    Message ID: 0x00000034
```

图 3-4 被动方发送的初始化消息和 KeepAlive 消息示例

```
⊟ Keep Alive Message
    0... .... = U bit: Unknown bit not set
    Message Type: Keep Alive Message (0x201)
    Message Length: 4
    Message ID: 0x0000001c
⊟ Address Message
    0... .... = U bit: Unknown bit not set
    Message Type: Address Message (0x300)
    Message Length: 22
    Message ID: 0x0000001d
  ⊟ Address List TLV
      00.. .... = TLV Unknown bits: Known TLV, do not Forward (0x00)
      TLV Type: Address List TLV (0x101)
      TLV Length: 14
      Address Family: IPv4 (1)
    ⊟ Addresses
        Address 1: 10.4.1.2
        Address 2: 5.5.5.5
        Address 3: 10.5.1.2
```

图 3-5 主动方发送的 KeepAlive 消息和地址消息示例

如果主动方 LSR_1 不能接受相关参数，则发送 Notification 消息给被动方 LSR_2，终止 LDP 会话的建立。

当双方都收到对端的 Keepalive 消息后，LDP 会话建立成功。但如果双方协商采用的是 DU 标签发布方式，则在主动端发了地址消息后，被动端也会专门发送一个地址消息给主动端。

3.1.4 LDP 的标签发布和管理

LDP 通过发送标签请求和标签映射消息，在 LDP 对等体之间通告 FEC 和标签的绑

定关系来建立 LSP，而标签的发布和管理由标签发布方式、标签分配控制方式和标签保持方式来决定。

1. 标签发布方式（Label Advertisement Mode）

在 MPLS 体系中，LSP 路径上的每个设备都会针对每个 FEC 从当前设备上**按由小到大的顺序**（最小标签为 1024）分配一个当前没有使用的入标签（可确保为每个 FEC 分配的标签都是唯一的）。但标签的分配总体来说是自下游向上游进行的，即先通过下游在本端设备上为对应的 FEC 分配出标签，然后本端设备再为该 FEC 分配入标签。本端设备为某 FEC 分配的出标签与下游节点为该 FEC 分配的入标签是相同的。**在同一设备上针对同一 FEC 的入标签与出标签可以相同，也可以不同，因为这两个标签是由相邻设备分别分配的。**

"标签发布方式"是指**是否要等到上游向自己发送某 FEC 的标签请求消息才向上游发送该 FEC 的标签映射消息，并为之分配出标签**，有如下两种方式。具有邻接关系的上、下游 LSR 必须对所使用的标签发布方式达成一致。

- 下游自主方式（DU，Downstream Unsolicited）：对于一个特定的 FEC，LSR 无需从上游 LSR 获得标签请求消息即可自主进行标签分配与分发。即不管是上游设备是否向本设备发出了标签请求，本设备在学习了新的 FEC 后可立即向上、下游对等体（**注意：会向所有对等体发送，不仅限向上游对等体发送**）发送该 FEC 的标签映射消息。

如图 3-6 所示，如果各 LSR 上配置的标签发布方式为 DU，则对于目的地址为 192.168.1.1/32 的 FEC，最下游（Egress）会通过标签映射消息主动向其上游（Transit）通告自己为主机路由 192.168.1.1/32 分配的入标签（将作为 Transit 的出标签）；然后 Transit 再利用标签映射消息主动向他的上游（Ingress）、下游（Egress）通告自己为主机路由 192.168.1.1/32 分配的入标签。**但向下游通告的标签映射消息最终不会起作用，因为下游已为该 FEC 分配好了入标签，且已建立好了该 FEC 的 LSP。**

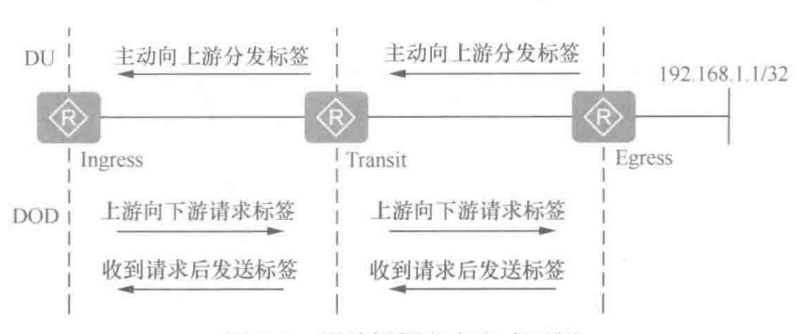


图 3-6 两种标签发布方式示例

【经验提示】DU 标签发布方式的最大优势就是简单（这也是华为设备上的缺省标签发布方式），无需上游设备请求，下游设备主动向上游设备分发标签，但这也同时带来了一个比较大的不足，那就是在一台设备上可能收到多个基于同一 FEC 的相同或不同的出标签，因为可能有多个对等体（也有到达同一目的主机的路由表项）向其分配基于同一 FEC 的出标签。这样一来，就会造成一个 FEC 映射了多个出标签，但同一时刻只有一个

标签映射是有效的。

■ 下游按需方式（DoD，Downstream on Demand）：对于一个特定的 FEC，LSR 只有在获得上游 LSR 发送的标签请求消息后才会向上游设备（**不会向下游设备**，因为在这种情形下，标签是严格按照从下游向上游方向分配的，只有上游设备才会向自己发送某 FEC 的标签请求消息）发送标签映射消息，进行标签分配。

如图 3-6 所示，如果各 LSR 上配置的标签发布方式为 DoD，对于目的地址为 192.168.1.1/32 的 FEC，如果最上游（Ingress）向其下游（Transit）发送标签请求消息，此时如果 Transit 还没有获取该 FEC 的出标签，则不会向 Ingress 发送分配标签的标签映射消息。但 Trasnit 可以向它的 Egress 发送标签请求消息，此时如果 Egress 以标签映射消息向 Transit 通告了 FEC 192.168.1.1/32 的入标签（将作为 Transit 的出标签），则 Transit 在为该 FEC 分配了入标签后即可通过标签映射消息向它的 Ingress 通告 192.168.1.1/32 的入标签（将作为 Ingress 的出标签）了。

【经验提示】DoD 标签发布方式虽然在节点向下游节点请求标签时可能会带来一些延时，但可以真正按需获取每个 FEC 的标签，使得各 LSR 上不会出现太多无用的标签映射。因为在 DoD 方式下，上游设备可只根据需要向一个下游设备请求标签，这样即使有多个对等体可以到达同一目的主机，其他对等体也不会向本地设备为此 FEC 分配标签。

2. 标签分配控制方式（Label Distribution Control Mode）

"标签分配控制方式"是指**是否要等到下游向自己发送了某 FEC 的标签映射消息才为该 FEC 分配入标签，并向上游发送该 FEC 的标签映射消息**，也有如下两种方式。

■ 独立标签分配控制方式（Independent）：本地 LSR 可以自主地分配一个入标签绑定到某个 FEC，然后向上游 LSR 进行标签通告，为上游 LSR 分配对应 FEC 的出标签，而无需等待下游 LSR 给本地 LSR 分配该 FEC 的出标签。

在这种分配控制方式下，LSR 在路由表中发现一个路由（对应一个 FEC）后，就会马上为该 FEC 分配一个标签，然后向上游 LSR 进行通告，根本不考虑其下游 LSR 是否已为该 FEC 分配了标签。这样就很可能会因为下游 LSR 还没有为该 FEC 分配标签、没有成功建立该 FEC 的 LSP，使其上游 LSR 即使已为该 FEC 分配了标签、建立 LSP，也无法与目的主机通信，造成数据丢失。

■ 有序标签分配控制方式（Ordered）：对于 LSR 上某个 FEC 的标签映射，只有当该 LSR 已经从其下一跳收到了基于此 FEC 的标签映射消息，或者该 LSR 就是此 FEC 的出节点时，该 LSR 才可以为此 FEC 分配入标签，然后向上游 LSR 发送此 FEC 的标签映射。

在这种分配控制方式下，LSR 必须要等到下游 LSR 已为本地 LSR 分配了某 FEC 的出标签后才能再为该 FEC 分配入标签。很显然，在这种分配控制方式中，最初进行入标签分配的是 Egress（出节点），Egress 的入标签也是作为倒数第二跳 Transit 的出标签，然后一级一级、有序地向上游进行标签分配。

标签分配控制方式与标签发布方式可以按照表 3-1 进行组合。

表 3-1 标签分配控制方式和标签发布方式的组合

标签发布方式 标签分配控制方式	下游自主方式 DU（Downstream Unsolicited）	下游按需方式 DoD（Downstream on Demand）
独立标签分配控制方式（Independent）	DU＋Independent：**两者都是独立方式**，LSR（Transit）无需等待下游为某 FEC 给本地 LSR 分配出标签，就会直接为该 FEC 分配入标签，然后向上游分配该 FEC 的出标签	DoD＋Independent：LSR（Transit）在收到上游发来的基于某 FEC 的标签请求消息后直接以标签映射消息进行回应，以本地为该 FEC 分配的入标签向为上游分配对应 FEC 的出标签，而不必等待下游给本地 LSR 为该 FEC 分配出标签
有序标签分配控制方式（Ordered）	DU＋Ordered：LSR（Transit）只要收到下游的标签映射消息，就可直接以本地为该 FEC 分配的入标签向上游分配出标签，不用等待上游发出标签请求消息	DoD＋Ordered：下游（Transit）在收到上游发送的基于某 FEC 的标签请求消息后，并且只有在收到下游发送的基于该 FEC 的标签映射消息后，才会以本地为该 FEC 分配的入标签向其上游分配出标签

3. 标签保持方式（Label Retention Mode）

标签保持方式是指 LSR 对收到的标签映射消息的处理方式，也有如表 3-2 所示的两种方式。LSR 收到的标签映射可能来自下一跳（本地对等体），也可能来自非下一跳（远端对等体）。

表 3-2 两种标签保持方式

标签保持方式	含义	说明
自由标签保持方式（Liberal）	对于从邻居 LSR 收到的标签映射，无论邻居 LSR 是不是自己的下一跳都保留	当网络拓扑变化引起下一跳邻居改变时： • 使用自由标签保持方式，LSR 可以直接利用原来非下一跳邻居发来的标签，迅速重建 LSP，但需要更多的内存和标签空间；
保守标签保持方式（Conservative）	对于从邻居 LSR 收到的标签映射，只有当邻居 LSR 是自己的下一跳时才保留	• 使用保守标签保持方式，LSR 只保留来自下一跳邻居的标签，节省了内存和标签空间，但 LSP 的重建会比较慢。保守标签保持方式通常与 DoD 方式一起，用于标签空间有限的 LSR

目前华为设备支持如下组合方式。

■ 下游自主方式（DU）＋ 有序标签分配控制方式（Ordered）＋ 自由标签保持方式（Liberal），该方式为缺省方式。**即 LSR 在收到下游标签映射后，可自主向其上游分配标签，且收到的标签全保留。**

■ 下游按需方式（DoD）＋ 有序标签分配控制方式（Ordered）＋ 保守标签保持方式（Conservative）。**即 LSR 在同时收到上游标签请求和下游标签映射后，才向上游分配标签，且只保留自己下一跳分配的标签。**

3.1.5 LDP LSP 的建立过程

LSP 的建立过程实际就是将 FEC 和标签进行绑定，并将这种绑定通告 LSP 上游相邻 LSR 的过程。但基于 LDP 的动态 LSP 建立过程相比第 2 章介绍的静态 LSP 建立过程要复杂许多，在此首先要了解 LDP LSP 建立的基本规则。

1. LDP LSP 建立的基本规则

LDP LSP 的建立是通过接收下游设备为 FEC 分配的出标签，或者同时为该 FEC 分配入标签，建立 FEC 与 MPLS 标签、出接口之间映射关系后而完成的。所以要建立基于某 FEC 的 LSP，首先就要为对应 FEC 分配标签。标签的分配必须遵循以下原则。

■ 入标签的分配是按由小到大（最小值为 1024）的顺序分配的，分配当前未分配的最小标签。

■ 同一链路上游设备为 FEC 分配的入标签一定要与下游设备为该 FEC 分配的出标签一致。

■ **在一台设备上针对同一 FEC 所分配的出标签可能有多个**（它们之间可以相同，也可以不同），分别来自不同下游对等体，也就是一个 FEC 可以映射多个出标签和出接口。

■ **在一台设备上针对同一 FEC 只会分配一个入标签**，即对于入标签，每个 FEC 在同一设备上都是唯一的。

每个路由表项都对应一个 FEC，缺省情况下，通过标签映射消息的通告，每个 FEC 都可能会在整个 MPLS 域网络的所有节点（包括本地设备）上建立 LSP，就像动态路由协议通过路由信息通告在整个网络或者特定区域内建立路由表项一样。

LDP LSP 建立的规则如下。

■ 在直接连接某 FEC 对应的网段（缺省仅为 32 位掩码的主机路由）的节点上会为该 FEC 仅创建一个包含入标签的 LSP（无出标签，也无入/出接口）。

■ 在其他节点上都会对非直连网段 FEC 同时创建两个 LSP：其中一个是用于指导从本地节点访问 FEC 所代表的目的主机的 LSP，仅包括出标签和出接口；另一个则是以本地节点为中间节点（Transit）的 LSP，用于指导上游设备访问 FEC 所代表的目的主机，同时包括入标签、出标签和出接口。

在如图 3-7 所示的示例中任意一节点上执行 **display mpls lsp** 命令，即可看到为本地所直连网段创建的那个仅包含入标签的 LSP。以下是在 AR2 上执行该命令的输出。

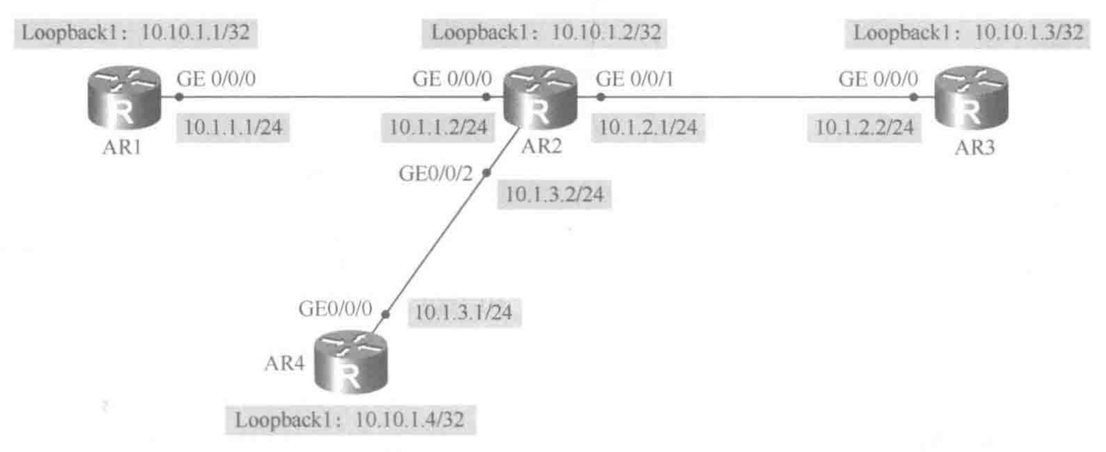


图 3-7 LDP LSP 建立示例一

```
<AR2>display mpls lsp
-------------------------------------------------------------------------------
                 LSP Information: LDP LSP
-------------------------------------------------------------------------------
FEC              In/Out Label  In/Out IF                      Vrf Name
10.10.1.4/32     NULL/3        -/GE0/0/2
10.10.1.4/32     1024/3        -/GE0/0/2
10.10.1.2/32     3/NULL        -/-
10.10.1.3/32     NULL/3        -/GE0/0/1
10.10.1.3/32     1025/3        -/GE0/0/1
10.10.1.1/32     NULL/3        -/GE0/0/0
10.10.1.1/32     1026/3        -/GE0/0/0
```

AR2 本地直连的 10.10.1.2/32 主机路由对应的那条 LSP 如以上输出信息中的粗体字部分。其他 LSP 均为到达非本地直连 32 位掩码网段所创建的 LSP，各有两条。如为 AR3 上的 10.10.1.3/32 这个主机路由就建立了以下两条 LSP。

```
FEC              In/Out Label  In/Out IF                      Vrf Name
10.10.1.3/32     NULL/3        -/GE0/0/1
10.10.1.3/32     1025/3        -/GE0/0/1
```

其中上面那条可以看成是把 AR2 当成到达 10.10.1.3/32 的 Ingress 的 LSP，即作为从 AR2 本地访问 10.10.1.3/32 的 LSP，它只包括出标签（无入标签）和出接口 GE0/0/1。下面那条则可以看成是把 AR2 当成到达 10.10.1.3/32 的 Transit 的 LSP，为其上游设备访问 10.10.1.3/32 的 LSP，它同时包括本地为 10.10.1.3/32 分配的入标签（1025）和下游 AR3 为它分配的出标签（3）和出接口 GE0/0/1。**其实这两条 LSP 的路径是相同的，不同的只是在 Ingress 上也为该 FEC 分配一个没有意义的入标签。**

以上这种现象与在其他节点上执行 **display mpls lsp** 命令的结果类似，都会为本地直连网段建立一个仅包括入标签（无出标签和出接口）的 LSP，为网络中的其他所有网段各建立两条 LSP（其中一条仅包括出标签和出接口，另一条则同时包括入/出标签和出接口），这两条 LSP 的出接口是一样的。下面是在图 3-7 中 AR1 上执行 **display mpls lsp** 命令的输出。

```
<AR1>display mpls lsp
-------------------------------------------------------------------------------
                 LSP Information: LDP LSP
-------------------------------------------------------------------------------
FEC              In/Out Label  In/Out IF                      Vrf Name
10.10.1.2/32     NULL/3        -/GE0/0/0
10.10.1.2/32     1024/3        -/GE0/0/0
10.10.1.3/32     NULL/1025     -/GE0/0/0
10.10.1.3/32     1025/1025     -/GE0/0/0
10.10.1.4/32     NULL/1024     -/GE0/0/0
10.10.1.4/32     1026/1024     -/GE0/0/0
10.10.1.1/32     3/NULL        -/-
```

以上示例中的 AR1 是到达 10.10.1.2/32、10.10.1.3/32 和 10.10.1.4/32 目的主机的 Ingress，没有上游设备了，所以在以上基于这些 FEC 建立的、同时带有入标签的 LSP 没有实际意义，参见以上输出信息中的粗体字部分。

2. LDP LSP 建立过程示例

下面以图 3-8 为例，并且以下游“自主标签发布方式”（无需上游请求）和“有序标签控制方式”（必须得到下游分配的标签）的组合介绍 LDP LSP 建立的主要步骤。

（1）当网络的路由改变时，假设 Egress 发现自己的路由表中出现了新的主机路由（如本示例中的 3.3.3.3/32），并且这一路由不属于任何现有的 FEC。

根据前面的介绍，Egress 会首先为路由表项新建一个 FEC，分配一个入标签（Egress 上通常是分配可以弹出的标签 3），建立 FEC 与入标签的映射（即 3.3.3.3→3），然后在本地建立一条该 FEC 的 LSP。

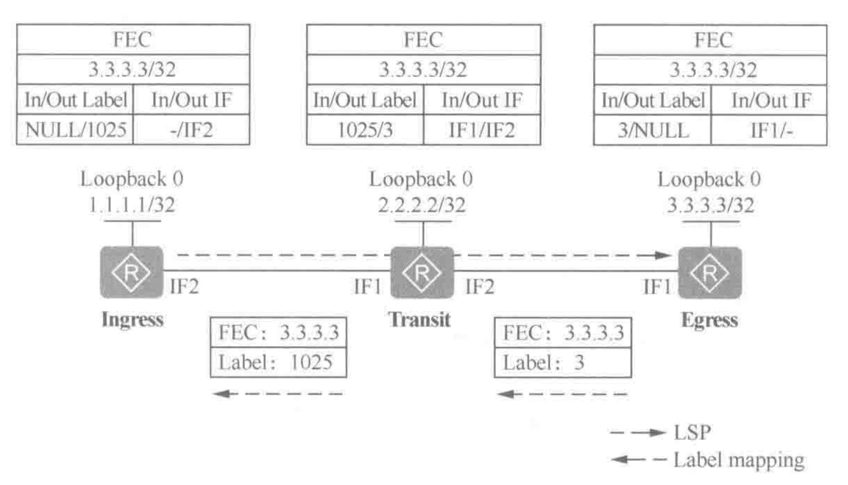


图 3-8　LDP LSP 建立示例二

在 Egress 上为 3.3.3.3/32 创建的本地 LSP 为：

```
FEC             In/Out Label   In/Out IF                Vrf Name
3.3.3.3/32      3/NULLL         -/-
```

（2）随后，Egress 会主动向其上游 Transit 发送标签映射消息，标签映射消息中包含为该 FEC 分配的入标签（3）和绑定的 FEC 等信息。

（3）Transit 收到标签映射消息后，判断标签映射的发送者（Egress）是否为该 FEC 的下一跳。若是，则在标签转发表中直接增加相应的转发条目，然后创建一个用于从本地访问 3.3.3.3/32 的 LSP，仅包括出 Egress 分配的出标签（3）和出接口。

因为 Transit 还有上游设备，所以还需要继续为上游设备分配该 FEC 的出标签。于是 Transit 又为该 FEC 分配一个入标签（1025），并在其标签转发表中增加相应的转发条目。因为本示例采用的是 DU 标签发布方式，所以会主动向上游 LSR（Ingress）发送基于该 FEC 的标签映射消息（3.3.3.3→1025），建立一条用于指导上游设备访问 3.3.3.3/32 网段的 LSP，此 LSP 的入标签和出标签分别为 1025 和 3。

在 Transit 上为 3.3.3.3/32 创建的两条 LSP 为：

```
FEC             In/Out Label   In/Out IF                Vrf Name
3.3.3.3/32      NULL/3          -/IF2
3.3.3.3/32      1025/3          -/IF2
```

（4）Ingress 收到标签映射消息后，判断标签映射的发送者（Transit）是否为该 FEC 的下一跳。若是，则在标签转发表中直接增加相应的转发条目，然后创建一个用于从本地访问 3.3.3.3/32 的 LSP，仅包括出 Transit 分配的出标签（1025）和出接口，如下所示。

```
FEC             In/Out Label   In/Out IF                Vrf Name
3.3.3.3/32      NULL/1025         -/IF2
3.3.3.3/32      1026/1025         -/IF2
```

注意 虽然 Ingress 节点后面无上游节点了，但它仍会再为该 FEC 分配一个入标签，创建一条同时包括入/出标签、出接口的 LSP，但实际上这条 LSP 是没有意义的，因为它上面没有上游设备了。

通过以上步骤就完成了整个 MPLS 网络中各节点基于 3.3.3.3/32 的各条 LSP 的建立，接下来各节点就可以进行该 FEC 对应的数据报文基于 MPLS 标签的转发了。

3. 代理 Egress

上面介绍的是普通 LDP LSP 的建立，此时 Egress 触发建立的 LSP 都是基于 Egress 本地所连接的 IP 网段（即本地路由）。如果 IP 网段不是 Egress 直接连接的，那么如果为这些网段触发建立 LSP 呢？此时要用到一种称之为代理 Egress（Proxy Egress）的功能。

代理 Egress 能够针对**非本地路由**触发建立 LSP 的 Egress 节点。当路由器使能倒数第二跳弹出时，倒数第二跳节点实际上就是一种特殊的代理 Egress，因为此时它所发送的报文中是不带有 MPLS 标签的。

一般情况下，代理 Egress 由配置产生，可应用于网络中有不支持 MPLS 特性的路由器场景（如在纯 IP 网络中），也可用于解决 BGP 路由负载分担问题。但一般来说，仅需建立 MPLS/IP 骨干网 LER 上直接连接的公网网段的 LSP 即可，无需建立非直连的 IP 网络中的网段 LSP。因为 LSP 的作用就是仅用于指导报文在 MPLS/IP 骨干网内的转发，在 IP 网络中的转发可直接依据 IP 路由进行。

如图 3-9 所示，LSR_1、LSR_2 和 LSR_3 处于同一个 MPLS 域中，LSR_4 未使能 MPLS LDP（在纯 IP 网络中）。此时，如果将建立 LSP 的策略配置为所有 IGP 路由都触发建立 LDP LSP，那么 LSR_3 将成为代理 Egress，使得 LSR_1、LSR_2 和 LSR_3 仍可以建立到 LSR_4 的 LDP LSP，当然更能建立到达 LSR_3 所连接的网段的 LDP LSP，即此时，LSR_3 既是本地网段的 Egress，又是非本地网段的代理 Egress。

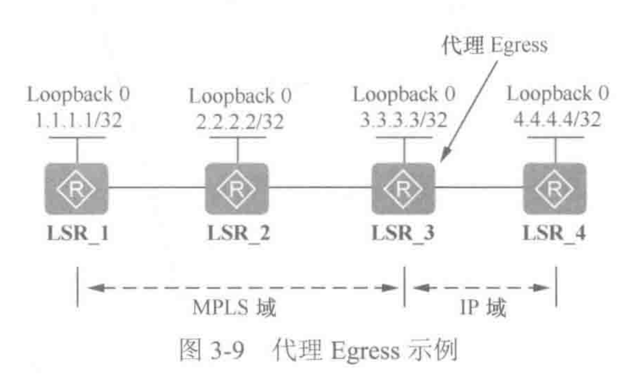


图 3-9　代理 Egress 示例

3.2 LDP 必选基本功能配置与管理

只有配置了 MPLS LDP 基本功能，才能组建 MPLS LDP 网络。在配置 MPLS LDP

基本功能之前，需配置静态路由或 IGP，使各节点间的 IP 路由可达。因为上游在接收到下游发来的标签映射消息时要根据路由表或转发表验证下一跳的合法性，还要根据路由表或转发表向上游进行标签映射消息发送。

LDP 基本功能所涉及的配置任务比较多，有必选的，也有许多只是针对一些特定应用场景，或者特定应用需求需要配置的可选配置任务。本节先介绍必选配置任务的具体配置方法，可选配置任务将在 3.3 节具体介绍。

3.2.1　配置 LDP 必选基本功能

LDP 基本功能中所包括的必须配置任务有以下几项（必须按顺序配置）。

（1）配置 LSR ID

LSR ID 用来在网络中唯一标识一个 LSR。**LSR 没有缺省的 LSR ID，必须手工配置。**为了提高网络的可靠性，推荐使用 LSR 某个 Loopback 接口的地址作为 LSR ID。需在 MPLS 域的所有节点上进行配置。

（2）使能全局 MPLS

只有使能了全局 MPLS，才可以配置 MPLS 的其他配置。需在 MPLS 域的所有节点上进行配置。

说明　前面两项配置任务的配置方法其实与静态 LSP 的对应配置是一样的。

（3）使能全局 MPLS LDP

只有使能了全局 MPLS LDP，才可以配置 MPLS LDP 的其他配置。需在 MPLS 域的所有节点上进行配置。

（4）配置 LDP 会话

配置 MPLS LDP 会话有以下方式。

■ 配置本地 LDP 会话

通常情况下，部署 MPLS LDP 业务时，需要配置本地 LDP 会话。

■ 配置远端 LDP 会话

远端 LDP 会话主要在不相邻的 LSR 之间建立，主要应用配置 Martini 方式的 VLL（一种 MPLS L2VPN，参见《华为 MPLS VPN 学习指南》一书）中，以构建 MPLS VPN 隧道。

本地 LDP 会话和远端 LDP 会话可以共存，即两个 LSR 之间既可以建立本地 LDP 会话，又可以建立远端 LDP 会话。在这种情况下，对于本地 LDP 会话和远端 LDP 会话进行两者都支持的相关配置时（如各种定时器、LDP 传输地址等的配置），两者的配置需要保持一致。

以上配置任务的具体配置步骤见表 3-3，除远端 LDP 会话外，其他各项任务均需要在各 LSR 上配置。

表 3-3　**配置 LDP 必选基本功能的步骤**

步骤	命令	说明
1	**system-view** 例如：<Huawei> **system-view**	进入系统视图

（续表）

步骤	命令	说明
		配置 LSD
2	**mpls lsr-id** *lsr-id* 例如：[Huawei] **mpls lsr-id** 1.1.1.1	配置本节点的 LSR ID，用于唯一标识一个 LSR，点分十进制格式（与 IPv4 地址格式一样，类似于 OSPF、BGP 路由器 ID）。 在网络中部署 MPLS 业务时，必须首先配置 LSR ID。**LSR 没有缺省的 LSR ID，必须手工配置**。为了提高网络的可靠性，推荐使用 LSR 某个 Loopback 接口的地址作为 LSR ID。 缺省情况下，没有配置 LSR ID，可用 **undo mpls lsr-id** 命令删除 LSR 的 ID。但如果要修改已经配置的 LSR ID，必须先在系统视图下执行 **undo mpls** 命令，然后再使用本命令配置
		使能全局 MPLS
3	**mpls** 例如：[Huawei] **mpls**	全局使能本节点的 MPLS，并进入 MPLS 视图。 缺省情况下，节点的 MPLS 能力处于未使能状态，可用 **undo mpls** 命令去使能全局 MPLS 功能，删除所有 MPLS 配置（LSR ID 除外）
4	**quit** 例如：[Huawei] **quit**	返回系统视图
		使能全局 MPLS LDP
5	**mpls ldp** 例如：[Huawei] **mpls ldp**	使能全局的 LDP 功能，并进入 MPLS-LDP 视图。 缺省情况下，没有使能全局的 LDP 功能，可用 **undo mpls ldp** 命令去使能全局 LDP 功能，删除所有 LDP 配置
6	**lsr-id** *lsr-id* 例如：[Huawei-mpls-ldp] **lsr-id** 2.2.2.3	（可选）配置 LDP 实例的 LSR ID（**仅适用于 LDP 实例**），点分十进制格式。在某些使用 VPN 实例的组网方案中（例如 BGP/MPLS IP VPN），如果 VPN 私网地址与 LSR ID 重叠，则需要为 LDP 另外配置 LSR ID，以保证 TCP 连接能够正常建立。 缺省情况下，LDP 实例的 LSR ID 等于节点的 LSR ID，可用 **undo lsr-id** 命令恢复缺省配置。推荐采用缺省值，修改和删除 LDP 实例的 LSR ID 会导致该实例下的所有会话被重建
		配置本地 LDP 会话
7	**quit** 例如：[Huawei-mpls-ldp] **quit**	返回系统视图
8	**interface** *interface-type interface-number* 例如：[Huawei] **interface** gigabitethernet 1/0/0	进入需要建立 LDP 会话的公网接口视图，**必须是三层接口**
9	**mpls** 例如：[Huawei-GigabitEthernet 1/0/0] **mpls**	使能以上接口的 MPLS 功能。 缺省情况下，接口的 MPLS 能力处于未使能状态，可用 **undo mpls** 命令去使能接口的 MPLS 功能，删除所在接口的 MPLS 配置
10	**mpls ldp** 例如：[Huawei-GigabitEthernet 1/0/0] **mpls ldp**	使能接口的 MPLS LDP 功能。 缺省情况下，接口的 MPLS LDP 能力处于未使能状态，可用 **undo mpls ldp** 命令去使能接口上的 MPLS LDP 功能

（续表）

步骤	命令	说明
	配置远端 LDP 会话（可与本地 LDP 会话同时配置）	
11	**quit** 例如：[Huawei-GigabitEthernet 1/0/0] **quit**	返回系统视图
12	**mpls ldp remote-peer** *remote-peer-name* 例如：[Huawei] **mpls ldp remote-peer** HuNan	创建 MPLS LDP 远端对等体，并进入 MPLS LDP 远端对等体视图。参数 *remote-peer-name* 指定远端对等体名称，字符串形式，不支持空格，**不区分大小写**，长度范围是 1～32。当输入的字符串两端使用双引号时，可在字符串中输入空格。 缺省情况下，没有创建远端对等体，可用 **undo mpls ldp remote-peer** *remote-peer-name* 命令删除远端对等体
13	**remote-ip** *ip-address* 例如：[Huawei-mpls-ldp-remote-rtc] **remote-ip** 10.1.1.1	配置 MPLS LDP 远端对等体的 IP 地址。**配置的远端对等体的 IP 地址必须是远端对等体的 LSR ID**。当 LDP LSR ID 和 MPLS LSR ID 不一致时，此处指的是 LDP LSR ID。修改或删除已经配置的远端对等体地址会导致相应的远端 LDP 会话被删除，造成 MPLS 业务被中断。 缺省情况下，没有配置 LDP 远端对等体的 IP 地址，可用 **undo remote-ip** 命令删除配置

LDP 会话配置好后，可用 **display mpls ldp session** [*peer-id* | [**all**] [**verbose**]]命令查看指定或所有对等体间的 LDP 会话状态，如果建立成功则显示为“Operational”状态。如下是一个执行该命令的输出示例。

```
<Huawei> display mpls ldp session verbose

 LDP Session(s) in Public Network
 ------------------------------------------------------------------------
 Peer LDP ID          : 2.2.2.2:0              Local LDP ID       : 1.1.1.1:0
 TCP Connection       : 1.1.1.1 <- 2.2.2.2
 Session State        : Operational            Session Role       : Passive
 Session FT Flag      : Off                    MD5 Flag           : Off
 Reconnect Timer      : ---                    Recovery Timer     : ---
 Keychain Name        : kc1

 Negotiated Keepalive Hold Timer    : 45 Sec
 Configured Keepalive Send Timer    : 3 Sec
 Keepalive Message Sent/Rcvd        : 438/438 (Message Count)
 Label Advertisement Mode           : Downstream Unsolicited
 Label Resource Status(Peer/Local) : Available/Available
 Session Age                        : 0000:01:49 (DDDD:HH:MM)
 Session Deletion Status            : No

 Capability:
  Capability-Announcement           : On
  mLDP P2MP Capability              : Off
  mLDP MBB Capability               : Off

 Outbound&Inbound Policies applied :
 outbound peer all split-horizon
```

```
Addresses received from peer: (Count: 3)
10.1.1.2              2.2.2.2              10.1.2.1
------------------------------------------------------------------------------
```

3.2.2 LDP 维护和管理命令

本节先集中介绍关于 LDP 基本功能（包括 3.3 节介绍的可选基本功能）配置和维护有关的命令，以便大家可以更好地理解后面所介绍的配置示例中的各种配置结果验证。

- **display default-parameter mpls management**：查看 MPLS 管理的缺省配置。
- **display default-parameter mpls ldp**：查看 MPLS LDP 的缺省配置。
- **display mpls interface** [*interface-type interface-number*] [**verbose**]：查看指定或所有使能了 MPLS 功能的接口信息。
- **display mpls ldp** [**all**] [**verbose**]：查看 LDP 的配置信息。
- **display mpls ldp interface** [*interface-type interface-number* | [**all**] [**verbose**]]：查看指定或所有使能了 LDP 功能的接口信息。
- **display mpls ldp adjacency** [**interface** *interface-type interface-number* | **remote**] [**peer** *peer-id*] [**verbose**]：查看指定或所有 LDP 邻接体信息。
- **display mpls ldp adjacency statistics**：查看 LDP 邻接体的统计信息。
- **display mpls ldp session** [[**all**] [**verbose**] | *peer-id*]：查看 LDP 会话状态信息。
- **display mpls ldp session statistics**：查看 LDP 对等体间的会话个数统计信息。
- **display mpls ldp peer** [[**all**] [**verbose**] | *peer-id*]：查看 LDP 会话的对等体信息。
- **display mpls ldp peer statistics**：查看 LDP 对等体个数的统计信息。
- **display mpls ldp remote-peer** [*remote-peer-name* | **peer-id** *lsr-id*]：查看指定或所有 LDP 远端会话的对等体信息。
- **display mpls ldp lsp** [all | *destination-address mask-length*] [**peer** *peer-id*]：查看指定或所有 LDP LSP 的建立信息。
- **display mpls ldp lsp statistics**：查看 LDP LSP 的统计信息。
- **display mpls route-state** [{ **exclude** | **include** } { **idle** | **ready** | **settingup** } * | *destination-address mask-length*] [**verbose**]：查看指定或所有动态 LSP 对应的路由相关信息。
- **display mpls lsp** [**verbose**]：查看 LSP 的建立信息。
- **display mpls lsp statistics**：查看当前处于 Up 状态的 LSP 数目，并显示 Ingress 节点、Transit 节点和 Egress 节点的当前激活的 LSP 数目。
- **display mpls label all summary**：查看 MPLS 所有标签的分配信息。

3.2.3 LDP 本地会话配置示例

如图 3-10 所示，LSRA、LSRC 为 IP/MPLS 骨干网的 PE 设备。LSRA 和 LSRC 上需要部署 MPLS L2VPN 或 L3VPN 业务来实现 VPN 站点的互联，因此 LSR 间需要配置本地 LDP 会话来建立 LDP LSP，实现承载 VPN 业务。

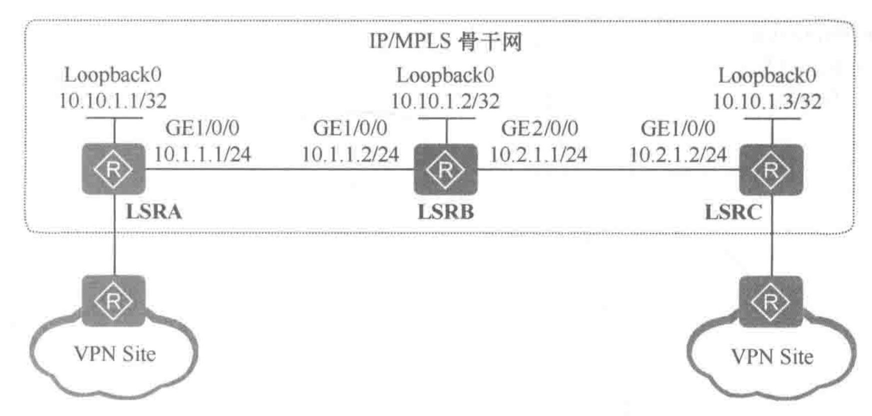


图 3-10　LDP 本地会话配置示例的拓扑结构

1. 基本配置思路分析

根据 3.2.1 节介绍，LDP 会话功能的配置是 LDP 必选基本功能的最后一项配置任务，所以必须先完成配置 LSR ID、全局使能 MPLS、使能全局 MPLS LDP 三项配置任务。另外，在配置 LDP 前，必须确保 MPLS 域中各 LSR 的路由互通，本示例假设采用 OSPF 路由协议来实现。

根据以上分析可得出本示例的基本配置思路如下。

（1）配置各 LSR 的接口（包括 Loopback 接口）IP 地址。

（2）配置各 LSR 的 OSPF 路由，实现骨干网的 IP 连通性。

（3）利用各 LSR 的 Loopback 接口配置各自的 LSR ID，全局使能 MPLS 和 LDP 功能。

（4）在各 LSR 间相连的接口上配置 LDP 本地会话。

2. 具体配置步骤

（1）配置各 LSR 各接口（包括 Loopback 接口）的 IP 地址。

LSRA 上的配置。

```
<Huawei> system-view
[Huawei] sysname LSRA
[LSRA] interface loopback 0
[LSRA-LoopBack0] ip address 10.10.1.1 32
[LSRA-LoopBack0] quit
[LSRA] interface gigabitethernet 1/0/0
[LSRA-GigabitEthernet1/0/0] ip address 10.1.1.1 24
[LSRA-GigabitEthernet1/0/0] quit
```

LSRB 上的配置。

```
<Huawei> system-view
[Huawei] sysname LSRB
[LSRB] interface loopback 0
[LSRB-LoopBack0] ip address 10.10.1.2 32
[LSRB-LoopBack0] quit
[LSRB] interface gigabitethernet 1/0/0
[LSRB-GigabitEthernet1/0/0] ip address 10.1.1.2 24
[LSRB-GigabitEthernet1/0/0] quit
[LSRB] interface gigabitethernet 2/0/0
[LSRB-GigabitEthernet2/0/0] ip address 10.2.1.1 24
[LSRB-GigabitEthernet2/0/0] quit
```

LSRC 上的配置。

```
<Huawei> system-view
[Huawei] sysname LSRC
[LSRC] interface loopback 0
[LSRC-LoopBack0] ip address 10.10.1.3 32
[LSRC-LoopBack0] quit
[LSRC] interface gigabitethernet 1/0/0
[LSRC-GigabitEthernet1/0/0] ip address 10.2.1.2 24
[LSRC-GigabitEthernet1/0/0] quit
```

（2）配置 OSPF 协议路由（包括 Loopback 接口主机路由），采用缺省 OSPF 路由进程 1，区域 0。

LSRA 上的配置。

```
[LSRA] ospf 1
[LSRA-ospf-1] area 0
[LSRA-ospf-1-area-0.0.0.0] network 10.10.1.1 0.0.0.0
[LSRA-ospf-1-area-0.0.0.0] network 10.1.1.0 0.0.0.255
[LSRA-ospf-1-area-0.0.0.0] quit
[LSRA-ospf-1] quit
```

LSRB 上的配置。

```
[LSRB] ospf 1
[LSRB-ospf-1] area 0
[LSRB-ospf-1-area-0.0.0.0] network 10.10.1.2 0.0.0.0
[LSRB-ospf-1-area-0.0.0.0] network 10.1.1.0 0.0.0.255
[LSRB-ospf-1-area-0.0.0.0] network 10.2.1.0 0.0.0.255
[LSRB-ospf-1-area-0.0.0.0] quit
[LSRB-ospf-1] quit
```

LSRC 上的配置。

```
[LSRC] ospf 1
[LSRC-ospf-1] area 0
[LSRC-ospf-1-area-0.0.0.0] network 10.10.1.3 0.0.0.0
[LSRC-ospf-1-area-0.0.0.0] network 10.2.1.0 0.0.0.255
[LSRC-ospf-1-area-0.0.0.0] quit
[LSRC-ospf-1] quit
```

以上配置完成后，在各节点上执行 **display ip routing-table** 命令，可以看到相互之间都学到了彼此的路由。

（3）在各 LSR 上配置 LSR ID（以各设备上的 Loopback0 接口 IP 地址作为 LSR ID），使能全局的 MPLS 和 MPLS LDP 功能。

LSRA 上的配置。

```
[LSRA] mpls lsr-id 10.10.1.1
[LSRA] mpls
[LSRA-mpls] quit
[LSRA] mpls ldp
[LSRA-mpls-ldp] quit
```

LSRB 上的配置。

```
[LSRB] mpls lsr-id 10.10.1.2
[LSRB] mpls
[LSRB-mpls] quit
[LSRB] mpls ldp
[LSRB-mpls-ldp] quit
```

LSRC 上的配置。

```
[LSRC] mpls lsr-id 10.10.1.3
```

```
[LSRC] mpls
[LSRC-mpls] quit
[LSRC] mpls ldp
[LSRC-mpls-ldp] quit
```

（4）在各 LSR 配置 LDP 本地会话，其实就是在各 LSR 相连的公网接口上使能 MPLS 和 MPLS LDP 功能。

LSRA 上的配置。

```
[LSRA] interface gigabitethernet 1/0/0
[LSRA-GigabitEthernet1/0/0] mpls
[LSRA-GigabitEthernet1/0/0] mpls ldp
[LSRA-GigabitEthernet1/0/0] quit
```

LSRB 上的配置。

```
[LSRB] interface gigabitethernet 1/0/0
[LSRB-GigabitEthernet1/0/0] mpls
[LSRB-GigabitEthernet1/0/0] mpls ldp
[LSRB-GigabitEthernet1/0/0] quit
[LSRB] interface gigabitethernet 2/0/0
[LSRB-GigabitEthernet2/0/0] mpls
[LSRB-GigabitEthernet2/0/0] mpls ldp
[LSRB-GigabitEthernet2/0/0] quit
```

LSRC 上的配置。

```
[LSRC] interface gigabitethernet 1/0/0
[LSRC-GigabitEthernet1/0/0] mpls
[LSRC-GigabitEthernet1/0/0] mpls ldp
[LSRC-GigabitEthernet1/0/0] quit
```

3. 实验结果验证

以上配置全部完成后，在节点上执行 **display mpls ldp session** 命令可以看到，LSRA 和 LSRB、LSRB 和 LSRC 之间的本地 LDP 会话状态为"Operational"，表示会话建立成功。除了这种状态外，还有以下几种状态。

- NonExistent：表示 LDP 会话的最初状态。在此状态双方互相发送 Hello 消息，在收到 TCP 连接建立成功事件的触发后变为 Initialized 状态。
- Initialized：表示 LDP 会话处于初始化状态。
- Open Sent：表示 LDP 会话进入初始化状态后，主动方给被动方发送了 Initialized 消息，并等待对方的回应。
- Open Recv：表示 LDP 会话进入初始化状态后，并且当双方都收到了对方发送的 Keepalive 消息后，LDP 会话进入 Operational 状态。

以下是在 LSRA 上执行 **display mpls ldp session** 命令的输出示例。

```
[LSRA] display mpls ldp session

 LDP Session(s) in Public Network
 Codes: LAM(Label Advertisement Mode), SsnAge Unit(DDDD:HH:MM)
 A '*' before a session means the session is being deleted.
 ------------------------------------------------------------------------------
 PeerID             Status       LAM  SsnRole  SsnAge      KASent/Rcv
 ------------------------------------------------------------------------------
 10.10.1.2:0        Operational  DU   Passive  0000:00:22  91/91
 ------------------------------------------------------------------------------
 TOTAL: 1 session(s) Found.
```

输出信息中的“PeerID”表示对等体的 LDP 标识符，格式为<LSR ID>：<标签空间>；“LAM”列表示标签发布方式；“SsnRole”表示本端 LSR 在 LDP 会话中的角色：Active 为主动方，LSR ID 值较大的一方，Passive 为被动端，LSR ID 值较小的一方；“SsnAge”表示 LDP 会话建立至今的时间间隔；“KASent/Rcv”表示 LDP 会话发送和接收的 keepalive 消息数。

3.2.4 远端 LDP 会话配置示例

如图 3-11 所示，LSRA、LSRC 是 IP/MPLS 骨干网的 PE 设备。LSRA 和 LSRC 上需要部署 MPLS L2VPN 业务来实现 VPN 站点的二层互联，因此 LSRA 和 LSRC 之间需要配置远端 LDP 会话来实现 VC（虚电路）标签交换。

说明 MPLS L2VPN 包括多种类型，如在《华为 MPLS VPN 学习指南》一书中介绍的 VLL（Virtual Leased Line，虚拟租用线）、VPLS（Virtual Private LAN Service，虚拟私网局域网服务）、PWE3（Pseudo-Wire Emulation Edge to Edge，端到端伪线仿真），这些都需要在隧道两端构建 VC（虚电路）。

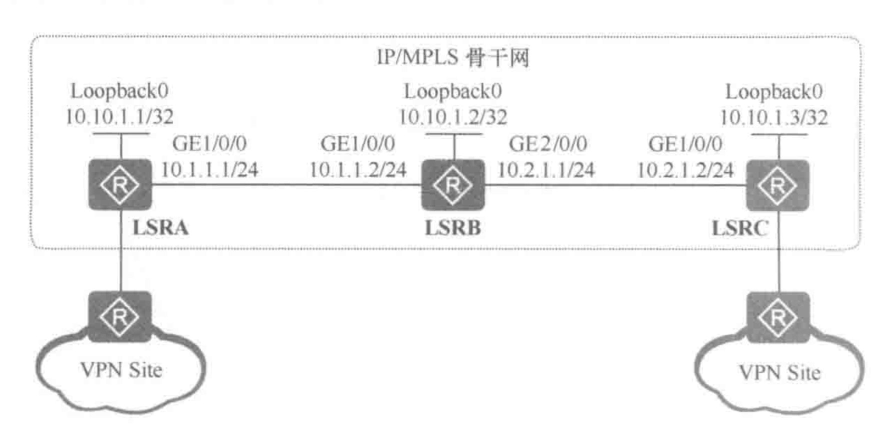


图 3-11 远端 LDP 会话配置示例的拓扑结构

1. 基本配置思路分析

在 MPLS L2VPN 中，隧道两端是 LER，如果两个 LER 是直接连接的，则在 LSR 上配置本地 LDP 会话后不仅可以建立 LDP LSP 来承载业务，还可以实现 VC 标签的交换。但本示例中 LSRA 和 LSRC 不是直连的，所以必须在它们之间配置远端 LDP 会话。

根据 3.2.1 节介绍的远端 LDP 会话配置步骤，可得出本示例的基本配置思路如下。

（1）配置各 LSR 的接口（包括 Loopback 接口）IP 地址。

（2）配置各 LSR 的 OSPF 路由，实现骨干网的 IP 连通性。

（3）在 LSRA 和 LSRC 上配置各自的 LSR ID，全局使能 MPLS 功能，使能全局 MPLS LDP 功能。

（4）在 LSRA 和 LSRC 上配置 LDP 远端会话，实现 VC 标签的交换。

说明 本示例虽然是在 LSRA 和 LSRC 之间建立 LDP 会话，但因为它们不是直连的，所

以前提也必须至少保证在LSRA和LSRC会话的路径上的各LSR的路由畅通，双向LSP建立成功，所以需要事先完成以上第（1）～（3）项配置任务。

2. 具体配置步骤

因为本示例与上节介绍的配置示例的拓扑结构和接口配置完全相同，故本示例中以上第（1）～（2）项配置任务的配置与3.2.3节介绍的示例的对应配置完全相同，参见即可。在第（3）项配置任务中，因为本示例仅需要在LSRA和LSRC之间建立远端LDP会话，无需与中间节点LSRB建立LDP会话，所以无需在LSRB上使能MPLS和LDP。在此具体列出以上第（3）和第（4）项配置任务的配置方法。

（3）在LSRA和LSRC上配置全局的MPLS和MPLS LDP能力。

\# LSRA上的配置

```
[LSRA] mpls lsr-id 10.10.1.1
[LSRA] mpls
[LSRA-mpls] quit
[LSRA] mpls ldp
[LSRA-mpls-ldp] quit
```

\# LSRC上配置

```
[LSRC] mpls lsr-id 10.10.1.3
[LSRC] mpls
[LSRC-mpls] quit
[LSRC] mpls ldp
[LSRC-mpls-ldp] quit
```

（4）在LSRA和LSRC之间配置远端LDP会话。要在两端LSR上创建远端对等体，然后指定远端对等体IP地址（为各自的Loopback接口的IP地址）。

\# LSRA上的配置

```
[LSRA] mpls ldp remote-peer lsrc
[LSRA-mpls-ldp-remote-lsrc] remote-ip 10.10.1.3
[LSRA-mpls-ldp-remote-lsrc] quit
```

\# LSRC上的配置

```
[LSRC] mpls ldp remote-peer lsra
[LSRC-mpls-ldp-remote-lsra] remote-ip 10.10.1.1
[LSRC-mpls-ldp-remote-lsra] quit
```

3. 配置结果验证

完成以上配置后，在LSRA或LSRC上执行**display mpls ldp session**命令可以看到，LSRA和LSRC之间的远端LDP会话状态为“Operational”，表示它们之间的远端LDP会话建立成功了，但并没有与LSRB（10.10.1.2）建立LDP会话。以下是在LSRA上执行该命令的输出示例（参见输出信息中的粗体字部分）。

```
[LSRA] display mpls ldp session

 LDP Session(s) in Public Network
 Codes: LAM(Label Advertisement Mode), SsnAge Unit(DDDD:HH:MM)
 A '*' before a session means the session is being deleted.
 ------------------------------------------------------------------------------
 PeerID             Status      LAM  SsnRole  SsnAge      KASent/Rcv
 ------------------------------------------------------------------------------
 10.10.1.3:0        Operational DU   Passive  0000:00:01  6/6
 ------------------------------------------------------------------------------
 TOTAL: 1 session(s) Found.
```

在远端 LDP 会话 LSRA、LSRC 上执行 **display mpls ldp remote-peer** 命令可以看到，LSR 的远端对等体的 LDP 会话参数配置信息，会话状态为 Active。以下是在 LSRA 上执行该命令的输出示例（参见输出信息中的粗体字部分）。

```
[LSRA] display mpls ldp remote-peer

                          LDP Remote Entity Information
 ------------------------------------------------------------------------------
 Remote Peer Name: lsrc
 Remote Peer IP      : 10.10.1.3              LDP ID         : 10.10.1.1:0
 Transport Address : 10.10.1.1              Entity Status : Active

 Configured Keepalive Hold Timer : 45 Sec
 Configured Keepalive Send Timer : ----
 Configured Hello Hold Timer         : 45 Sec
 Negotiated Hello Hold Timer         : ----
 Configured Hello Send Timer         : 15 Sec
 Configured Delay Timer               : 10 Sec
 Hello Packet sent/received         : 6347/6307
 Label Advertisement Mode            : Downstream Unsolicited
 Remote Peer Deletion Status       : No
 Auto-config                               : ---
 ------------------------------------------------------------------------------
 TOTAL: 1 Peer(s) Found.
```

3.3 配置 LDP 可选基本功能

在 LDP 的基本功能配置中，除了前面介绍的必选配置任务的配置外，还可根据实际的网络应用场景和需求选择配置以下任务：

- （可选）配置 LDP 传输地址；
- （可选）配置 LDP 会话的定时器；
- （可选）配置 PHP 特性；
- （可选）配置 LDP 标签分配控制方式；
- （可选）配置 LDP 标签发布方式；
- （可选）配置 LDP 自动触发 DoD 请求功能；
- （可选）配置 MPLS MTU；
- （可选）配置 MPLS 对 TTL 的处理；
- （可选）配置 LDP 标签策略；
- （可选）禁止向远端对等体分配标签；
- （可选）配置 LDP LSP 建立的触发策略；
- （可选）配置 Label Withdraw 消息延迟发送。

3.3.1 配置 LDP 传输地址和 PHP 特性

本节对 LDP 传输地址和 PHP 特性这两项可选配置任务的配置方法进行了集中介绍，但前提是必须在已完成了 3.2.1 节必选功能配置的基础上进行。

1. 配置 LDP 传输地址

LDP 传输地址就是用来在对等体间建立 LDP 对等间会话的 IP 地址，因为 LDP 会话是基于 TCP 连接的，当两台 LSR 之间要建立 LDP 会话前，必须先确认对端的 LDP 传输地址。通常情况下，这个 LDP 传输地址是无需另外配置的，因为缺省是使用 LSR ID（Loopback 接口地址）作为传输地址的。**但当本端 Loopback 接口的 IP 地址是公网 IP 地址，而对端 Loopback 接口的 IP 地址是私网 IP 地址时，则需要为本端也配置私网 IP 地址作为传输地址**，使对等体间能够使用私网 IP 地址建立连接。

配置 LDP 传输地址的方法很简单，在需要建立 LDP 会话的接口视图下通过 **mpls ldp transport-address** { *interface-type interface-number* | **interface** }命令进行。命令中的参数和选项说明如下。

■ *interface-type interface-number*：二选一参数，指定 LDP 使用此接口 IP 地址作为 TCP 传输地址。

■ **interface：二选一选项，指定 LDP 使用当前接口的 IP 地址作为 TCP 传输地址。**

缺省情况下，公网的 LDP 传输地址等于节点的 LSR ID，私网的传输地址等于启用了 MPLS LDP 功能的物理接口的主 IP 地址，可用 **undo mpls ldp transport-address** 命令恢复缺省配置。修改 LDP 传输地址的配置时，会话不会立刻被中断，而是等待 Hello 保持定时器超时后中断。

当两个 LSR 之间存在多条链路且要在多条链路上建立 LDP 会话时，**会话的同一端的接口都应采用默认的传输地址，或者配置相同的传输地址**。如果会话的一端接口配置了不同的传输地址，将导致 LDP 会话只能建立在一条链路上。

2. 配置 PHP 特性

PHP（Penultimate Hop Popping 倒数第二跳弹出）特性就是在倒数第二个节点上弹出标签的特性。因为在 LSP 的最后一跳节点（Egress），所以已不需要再进行标签交换。

通过在倒数第二跳节点上配置 PHP 特性，使倒数第二跳节点在向最后一跳节点发送报文时将最外层的出标签弹出（**如果最外层出标签被弹出后只剩下栈底标签，也将被弹出**），以使最后一跳可以直接进行 IP 转发或者下一层标签转发，在减少最后一跳标签交换的负担。但并不是在倒数第二跳配置了 PHP 特性就一定能将最外层标签弹出，还要视最后一跳原来为其分配的标签类型而定，因为并不是所有标签均支持被弹出。

在出节点（Egress）上配置 PHP 特性的方法是在 MPLS 视图下通过 **label advertise** { **explicit-null** | **implicit-null** | **non-null** }命令进行的，该命令是配置出节点向倒数第二跳分配的标签的类型，命令中的选项说明如下。

■ **explicit-null**：多选一选项，不支持 PHP 特性，指定出节点向倒数第二跳分配显式空标签，显式空标签的值为 0。如果出节点分配给倒数第二跳节点的标签值为 0，则倒数第二跳 LSR 需要将值为 0 的标签正常压入报文标签值顶部，转发给出节点。出节点发现报文携带的标签值为 0，则将标签弹出（**即标签的弹出是在出节点进行的，不是在倒数第二跳节点进行的**）。

■ **implicit-null**：多选一选项，支持 PHP 特性，指定出节点向倒数第二跳分配隐式空标签，隐式空标签的值为 3。倒数第二跳节点进行标签交换时，如果发现交换后的标签值为 3（**即标签的弹出是在倒数第二跳节点进行的**），则将标签弹出，并将报文发给最后一跳。最后一跳收到该报文直接进行 IP 转发或下一层标签转发。

■ **non-null**：多选一选项，不支持 PHP 特性，指定出节点向倒数第二跳正常分配标签，分配的标签值不小于 16。

缺省情况下，出节点向倒数第二跳分配隐式空标签（implicit-null），推荐采用缺省配置，可以减少出节点的转发压力，提高转发效率。



配置以上 **label advertise** 命令后，仅在以下情况配置才会生效。

■ 系统发生主备倒换。

■ 用户手工执行了以下操作。

- 执行 **reset mpls ldp** 命令重启 LDP 公网实例。
- 修改当前 LDP LSP 建立的触发策略，且修改的 LDP LSP 建立的触发策略范围是由小变大的。例如：LDP LSP 建立的触发策略由 **none** 修改为 **all**，则配置对所有的 LDP LSP 生效；LDP LSP 建立的触发策略由 **host** 修改为 **all**，则配置仅对除主机路由外的其他路由建立的 LDP LSP 生效。

配置 LDP LSP 建立的触发策略的方法将在本章后面介绍。

3.3.2 配置 LDP 会话的定时器

LDP 协议在建立和维护 LDP 会话的过程使用了表 3-4 所示的多种定时器，这些定时器一般情况下直接使用缺省配置即可。

表 3-4 LDP 会话使用的定时器

定时器	描述	使用建议
Hello 发送定时器，包括以下两种： • 链路 Hello 发送定时器（即本地 LDP 会话中的 Hello 发送定时器）； • 目标 Hello 发送定时器（即远端 LDP 会话中的 Hello 发送定时器）	LSR 使用 Hello 定时器周期性地发送 Hello 消息，向对等体 LSR 通告他在网络中的存在，并建立 Hello 邻接关系	在网络状况不是很好的网络中，可以适当调小 Hello 发送定时器的时间，以便尽早发现网络故障
Hello 保持定时器，包括以下两种： • 链路 Hello 保持定时器（即本地 LDP 会话中的 Hello 保持定时器）； • 目标 Hello 保持定时器（即远端 LDP 会话中的 Hello 保持定时器）	建立了 Hello 邻接关系的 LDP 对等体之间，通过周期性发送 Hello 报文表明自己希望继续维持这种邻接关系。如果在 Hello 保持定时器超时后，仍没有收到来自某对等体新的 Hello 报文，则拆除与该对等体之间的 Hello 邻接关系	在链路状态不稳定或者发送报文数量较大的网络中，可以适当调大 Hello 保持定时器的时间，以避免会话被频繁拆除和建立
Keepalive 发送定时器	LDP 会话建立以后，对等体间以 Keepalive 发送定时器为周期，周期性地向对端发送 Keepalive 消息，用于保持它们间的 LDP 会话	在网络状况不是很好的网络中，可以适当调小 Keepalive 发送定时器的时间，以便尽早发现网络故障
Keepalive 保持定时器	LDP 对等体之间通过 LDP 协议报文（PDU）维持 LDP 会话，如果在 Keepalive 保持定时器超时后，仍没有收到来自某对等体新的 LDP PDU，则关闭它们之间的连接，结束 LDP 会话	在链路状态不稳定的网络中，可以适当调大 Keepalive 保持定时器的时间，以尽量避免 LDP 会话振荡

（续表）

定时器	描述	使用建议
指数回退定时器	LDP 会话初始化消息处理失败或者收到对端 LSR 会话初始化消息的拒绝通知后，会话发起的主动端会启动指数回退定时器，定期尝试重新建立会话	当设备升级时，需要延长会话发起端尝试建立会话的周期，可以配置较大的指数回退定时器的初始值和最大值。当设备承载业务容易发生闪断时，需要缩短会话发起端尝试建立会话的周期，可以配置较小的指数回退定时器的初始值和最大值

从表 3-4 中可以看出 Hello 发送定时器和 Hello 保持定时器均有两种：

- 在本地 LDP 会话中使用的链路 Hello 发送定时器和链路 Hello 保持定时器；
- 在远端 LDP 会话中使用的目标 Hello 发送定时器和目标 Hello 保持定时器。

其实这两种 Hello 发送定时器和链路 Hello 保持定时器在两种会话中的作用是相同的，只是作用的 LDP 会话类型不同，且它们在参数值的取值范围和缺省值上也有所不同。后面三种定时器在两种 LDP 会话中的作用和参数取值范围、缺省值都是一样的。

以上这些 LDP 定时器的配置方法见表 3-5，但在本地 LDP 会话和远端 LDP 会话中，这些参数的配置对象不同：本地会话是在本地接口视图下进行的，而远端会话是在对等体视图下进行的，也必须是在已完成了 3.2.1 节必选功能配置的基础上进行。

注意 当一 LSR 本地和远端会话共存时，本地和远端会话的 Keepalive 发送定时器和 Keepalive 保持定时器必须保持一致。

表 3-5　　配置本地 LDP 会话定时器的步骤

步骤	命令	说明
1	**system-view** 例如：<Huawei> **system-view**	进入系统视图
2	**interface** *interface-type interface-number* 例如：[Huawei] **interface** gigabitethernet 1/0/0	（二选一）本地 LDP 会话时，进入建立 LDP 会话的接口视图
	mpls ldp remote-peer *remote-peer-name* 例如：[Huawei] **mpls ldp remote-peer** HuNan	（二选一）远端 LDP 会话时，进入 MPLS-LDP 远端对等体视图
3	**mpls ldp timer hello-send** *interval* 例如：[Huawei-GigabitEthernet1/0/0] **mpls ldp timer hello-send** 10 或 [Huawei-mpls-ldp-remote-huan] **mpls ldp timer hello-send** 10	配置链路或目标 Hello 发送定时器，整数形式，取值范围是 1～65535，单位是秒。 链路/目标 Hello 发送定时器的实际生效值=Min{链路/目标 Hello 发送定时器的配置值，链路/目标 Hello 保持定时器值的 1/3}。 缺省情况下，链路/目标 Hello 发送定时器的值是链路/目标 Hello 保持定时器值的 1/3，可用 **undo mpls ldp timer hello-send** 命令恢复缺省配置

（续表）

步骤	命令	说明
4	**mpls ldp timer hello-hold** *interval* 例如：[Huawei-GigabitEthernet1/0/0] **mpls ldp timer hello-send** 60 或 [Huawei-mpls-ldp-remote-huan] **mpls ldp timer hello-send** 60	配置链路/目标 Hello 保持定时器，整数形式，取值范围是 3～65535，单位是秒（65535 表示永不超时）。 本端 LSR 所配置的 Hello 保持定时器的值并不一定等于实际生效的定时器的值。**实际生效的定时器值等于会话两端 LSR 所配置的定时器的较小值，如果这个值小于 9，则 Hello 保持定时器的值等于 9。** 缺省情况下，链路 Hello 保持定时器的值是 15s，目标 Hello 保持定时器为 45s，可用 **undo mpls ldp timer hello-hold** 命令恢复缺省值
5	**mpls ldp timer keepalive-send** *interval* 例如：[Huawei-GigabitEthernet1/0/0] **mpls ldp timer keepalive-send** 10 或 [Huawei-mpls-ldp-remote-huan] **mpls ldp timer keepalive-send** 10	配置本地/远端 LDP 会话的 Keepalive 发送定时器，整数形式，取值范围是 1～65535，单位是秒。 本地/远端 LDP 会话 Keepalive 发送定时器的实际生效值=Min{本地/远端 LDP 会话 Keepalive 发送定时器的配置值，本地/远端 LDP 会话 Keepalive 保持定时器值的 1/3}。 【注意】如果两个 LSR 之间使能 LDP 的链路条数超过 1 条，所有链路的 Keepalive 发送定时器时间，以及下面一步将要配置的 Keepalive 保持定时器时间都必须相同，否则 LDP 会话可能不稳定。或者两个 LSR 之间的链路条数为 1 条，但既配置了本地会话又配置了远端会话，那么本地会话和远端会话的 KeepAlive 发送定时器时间和 Keepalive 保持定时器时间必须相同，否则 LDP 会话可能不稳定或者 LSP 会话无法建立。 缺省情况下，本地/远端 LDP 会话的 Keepalive 发送定时器的值是本地/远端 LDP 会话的 Keepalive 保持定时器值的 1/3，可用 **undo mpls ldp timer keepalive-send** 命令恢复缺省配置
6	**mpls ldp timer keepalive-hold** *interval* 例如：[Huawei-GigabitEthernet1/0/0] **mpls ldp timer keepalive-hold** 60 或 [Huawei-mpls-ldp-remote-huan] **mpls ldp timer keepalive-hold** 60	配置本地/远端 LDP 会话的 Keepalive 保持定时器，整数形式，取值范围是 30～65535，单位是秒。 本端/远端 LSR 所配置的 Keepalive 保持定时器的值并不一定等于实际生效的定时器的值。实际生效的定时器值等于 LDP 本地/远端会话两端 LSR 所配置的定时器的较小值。 缺省情况下，本地/远端 LDP 会话的 Keepalive 保持定时器的值是 45s，可用 **undo mpls ldp timer keepalive-hold** 命令恢复缺省值
7	**quit** 例如：[Huawei-GigabitEthernet1/0/0] **quit** 或 [Huawei-mpls-ldp-remote-huan] **quit**	返回系统视图
8	**mpls ldp** 例如：[Huawei] **mpls ldp**	进入 MPLS-LDP 视图

（续表）

步骤	命令	说明
9	**backoff timer** *init max* 例如：[Huawei-mpls-ldp] **backoff timer** 20 160	配置指数回退定时器，当会话再出现故障时，将按照此次配置的指数回退定时器的初始值和最大值来尝试重建会话。命令的参数说明如下。 • *Init*：指定指数回退定时器的初始值，整数形式，取值范围是 5～2147483，单位是秒。 • *max*：指定指数回退定时器的最大值，整数形式，取值范围是 5～2147483，单位是秒。 【说明】LDP 会话初始化消息处理失败或者收到对端 LSR 会话初始化消息的拒绝通知后，会话发起的主动端会启动指数回退定时器，定期尝试重新建立会话。指数回退定时器启动后，会话发起端第一次等待尝试重新建立会话的时间是指数回退定时器的初始值，随后每次的等待时间是前一次等待时间的 2 倍，直到等待时间达到了指数回退定时器的最大值，以后的等待时间均是指数回退定时器的最大值。 缺省情况下，指数回退定时器的初始值是 15s，最大值是 120s。建议配置指数回退定时器初始值不小于 15s，最大值不小于 120s

3.3.3　配置标签发布和分配控制方式

标签发布方式和标签分配控制方式均已在 3.1.4 节有详细介绍，它们都有缺省配置，所以一般情况下无需进行本节所介绍的配置。但在实际应用中，如果确实需要更改缺省配置，则可采取本节介绍的配置方法，前提也必须是已完成了 3.2.1 节必选功能配置。

1．配置 LDP 标签发布方式

LDP 标签发布方式是针对下游设备（按具体 LSP 方向）而言的，是下游设备向上游设备发送标签映射消息的方式，可以简单地理解为主动和被动两种方式，分别对应 3.1.4 中介绍的 DU 和 DoD 方式。

主动方式无需上游设备向其发出标签请求消息，下游设备都会主动向上游设备发送标签映射消息，为上游设备分配基于某 FEC 的出标签。这种方式可能使上游设备会收到很多在当前并不需要的标签，浪费了内存空间。被动方式也是只有上游设备主动向下游设备发出标签映射请求消息后，下游设备才会向上游设备发送标签映射消息，是真正的按需发布方式，这种方式可使上游设备只获取自己真正需要的标签，节省了设备的内存空间。

缺省情况下，下游设备会自主向上游设备主动发送标签映射消息（采用 DU 方式），当网络发生故障时，业务可以迅速切换到备份路径上，提高了网络的可靠性。但由于 MPLS 网络中边缘设备（LER，包括 Ingress 和 Egress）通常属于低端设备，当网络规模比较大时，为了保证网络的稳定，需要尽可能地减轻边缘设备的负担，如果它们的下游设备仍总是主动向它们发送标签映射消息，则可能造成边缘上有大量空闲、当前无用的标签（**因为并不是所有在标签信息表 LIB 中的标签都是当前有用的**），不仅会消耗设备的内存资源，还会在进行标签映射表项查找时仍消耗设备的系统资源。

在这种情况下，我们可以在边缘设备和它们的对等体上同时配置 LDP 标签发布方式为上游按需向下游请求标签（采用 DoD 方式），这样仅当下游设备在收到上游边缘设备的标签请求消息后才向这些边缘设备发送标签映射消息，以减少边缘设备 LIB 的大小，节省边缘设备的内存和系统资源。

本项配置任务需要在 MPLS 域两端 LER 与相邻 Transit 相连的两端接口上同时配置（**如果仅针对特定方向 LSP 时，只需在该 LSP 方向上 Ingress 出接口和相邻下游 Transit 的入接口上配置**），在具体接口视图下通过 **mpls ldp advertisement** { **dod** | **du** }命令配置即可，命令中的两个选项就是前面提到的两种 LDP 标签发布方式，可用 **undo mpls ldp advertisement** 命令恢复缺省设置。但修改标签发布方式会导致 LDP 会话重建，造成 MPLS 业务短时间中断。且当对等体之间存在多链路的时候，所有接口的标签发布方式必须相同。

【经验提示】这里之所以在两端的边缘设备和它们相邻的 Transit 设备相连接口上同时配置，是因为 MPLS LSP 隧道是单向的，而在一个网络中，通信方向是双向的，需要建立两条相反方向的 LSP 和 MPLS 隧道，一台边缘设备对于一条 LSP 是 Ingress（或 Egress），但对于另一 LSP 它又是 Egress（或 Ingress）。

当边缘设备作为某 LSP 的 Ingress 时，与其相邻的 Transit 就是其下游设备，为了减少边缘设备负担，就需要在 Transit 与边缘设备相连的接口上修改 LDP 标签发布方式；而当边缘设备作为某 LSP 的 Egress 时，与它相邻的 Transit 就是其上游设备，这时又需要在 Egress 连接相邻 Transit 的接口上修改标签发布方式。

2. 配置 LDP 标签分配控制方式

标签分配控制方式也是针对下游设备（按具体 LSP 方向）而言的，是指下游设备在向上游设备发布标签的时机或者条件，分为独立（Independent）方式和有序（Ordered）方式两种，具体可参见 3.1.4 节。

缺省情况下，LDP 的标签分配控制方式为有序标签分配控制（即采用 Ordered 方式），即只有当该 LSR 已经收到此 FEC 下一跳的标签映射消息，或者该 LSR 就是此 FEC 的出节点（Egress）时，该 LSR 才可以向上游 LSR 发送此 FEC 的标签映射。但在重新部署业务时，如果希望业务能够快速建立，则可以配置采用独立标签分配控制（即采用 Independent 方式）方式，即本地 LSR 可以自主地分配一个入标签绑定到某个 FEC，然后通告给上游 LSR，而无需等待下游 LSR 发来的标签映射消息。

针对特定 LSP 而言，本项配置任务只需在 Egress 和 Transit 的 MPLS LDP 视图下通过 **label distribution control-mode** { **independent** | **ordered** }命令进行配置，命令中的两个选项就是前面提到的两种 LDP 标签分配控制方式，可用 **undo label distribution control-mode** 命令恢复为缺省配置。如果要针对所有 LSP 来配置，则需要在所有 LSR 设备上配置，同样是因为这样任一边缘设备既可能是某些 LSP 的 Ingress，又可能是另一些 LSP 的 Egress。

3.3.4 配置 LDP 自动触发 DoD 请求功能

在 3.3.3 节已介绍到，当网络规模比较大时，为了尽可能地减轻边缘设备（LER）的负担，需要配置与边缘设备相邻的 Transit 入接口的标签发布方式为 DoD。但在跨 IGP 域（存

在多种路由类型）网络的远端LDP会话中，由于边缘设备之间可能无法学习到对方的精确路由（如不同IGP域中采用缺省路由或聚合路由进行相互通告），即使配置了跨域扩展功能也无法建立LDP LSP。此时可以在边缘设备上配置自动触发采用DoD的标签发布方式向下游请求指定的或者所有的远端对等体的标签映射消息，这样就可以建立LDP LSP了。

基于以上分析，如果仅针对单方向LSP而言，仅需在入口LER设备上配置，如果要针对双向LSP，则要同时在入/出口LER上配置，但在配置前需已完成以下配置。

■ 在两远端对等体间配置LDP远端会话（参见3.2.1节）。

■ 在两远端对等体间配置LDP跨域扩展（将在下章介绍），使能LDP按照最长匹配方式查找路由建立LSP。

■ 在LER与相邻Transit相连的接口上配置标签发布方式为DoD方式（参见3.3.3节）。

从以上可以看出，LDP自动触发DoD请求功能需要同时结合LDP远端会话、LDP跨域扩展、LDP DoD标签发布方式三种功能来实现。

完成以上配置后，可根据需要在两远端对等体上选择以下两种配置方法之一。

（1）配置自动触发采用DoD的标签发布方式向下游请求**所有的标签映射消息**。

这种请求方式是配置在采用DoD的标签发布方式下，自动向下游请求所有的远端对等体的标签映射消息，具体的配置方法是在MPLS LDP视图下通过 **remote-peer auto-dod-request** 命令进行。缺省情况下，没有配置在采用DoD的标签发布方式下自动向下游所有的远端对等体请求标签映射消息，可用 **undo remote-peer auto-dod-request** 命令恢复缺省配置。

（2）配置自动触发采用DoD的标签发布方式向下游请求**指定的远端对等体的标签映射消息**。

这种请求方式是配置在采用DoD的标签发布方式下，自动向下游指定的远端对等体（即仅针对特定的远端对等体）请求标签映射消息，具体的配置步骤见表3-6。

表3-6 配置向指定远端请求标签映射消息的步骤

步骤	命令	说明
1	**system-view** 例如：<Huawei> **system-view**	进入系统视图
2	**mpls ldp remote-peer** *remote-peer-name* 例如：[Huawei] **mpls ldp remote-peer** HuNan	进入MPLS LDP远端对等体视图，该远端对等体必须已在创建LDP远端会话时创建
3	**remote-ip auto-dod-request** [**block**] 例如：[Huawei-mpls-ldp-remote-hunan] **remote-ip auto-dod-request**	配置自动触发采用DoD的标签发布方式向下游请求以上远端对等体的标签映射消息。如果需要屏蔽自动触发采用DoD的标签发布方式向下游请求指定的远端对等体的标签映射消息，可以选择 **block** 可选项。 缺省情况下，继承全局的 **remote-peer auto-dod-request** 命令的配置属性，可用 **undo remote-ip auto-dod-request** 命令恢复缺省配置

3.3.5 LDP自动触发DoD请求功能配置示例

如图3-12所示，LSRA、LSRD是两台网络边缘设备。为了建立PW（伪线，L2VPN

中的概念，本书后面将有介绍），必须在 LSRA 和 LSRD 之间建立 LDP 远端会话，从而建立公网隧道。由于网络规模宏大，要求尽可能地节省网络资源，减少不必要的 LSP 和 MPLS 表项。

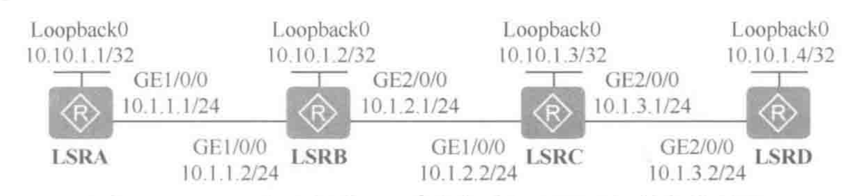


图 3-12 LDP 自动触发 DoD 请求功能配置示例的拓扑结构

1. 基本配置思路分析

根据本示例的要求，可以配置 LDP 自动触发 DoD 请求功能来实现。上节已分析到，这项功能要同时结合 LDP 远端会话、LDP 跨域扩展、LDP DoD 标签发布方式三种功能，所以需要事先完成这三项功能的配置。

在本示例中，为了模拟 LDP 跨域扩展功能所需的多 IGP 环境，现假设仅在 LSRB 与 LSRC 的连接中采用 IS-IS 路由，而 LSRA 与 LSRB、LSRC 与 LSRD 均采用静态路由，其中 LSRA 到 LSRB、LSRD 到 LSRC 均采用缺省路由。然后在 IS-IS 路由进程中引入静态路由。

这样做的目的是使 IS-IS 路由域中的 LSRB 和 LSRC 均只获得了 LSRA 和 LSRD 的缺省路由，无法向它们的上游设备发布基于 LSRA 或 LSRD 上精确路由的标签映射信息，即使 LSRA 和 LSRD 之间无法根据对端的精确路由建立各条 LSP，但可依据它们之间的 LDP 远端会话、LDP 跨域扩展功能提供的按照最长匹配方式查找路由建立 LSP 功能，以及 LDP 自动触发 DoD 请求功能来建立各条 LSP。

根据以上分析，可得出本示例的基本配置思路如下。

（1）配置各 LSR 的各接口（包括 Loopback 接口）的 IP 地址。

（2）在 LSRB 的 GE2/0/0 接口和 LSRC 的 GE1/0/0 接口上配置 IS-IS 路由，并配置 LSRA 到 LSRB、LSRD 到 LSRC 的缺省路由，以及 LSRB 到 LSRA、LSRC 到 LSRD 的静态路由。

（3）在各 LSR 上全局使能 MPLS 和 MPLS LDP 功能，在各公网接口（不包括 Loopback 接口）上使能 MPLS 和 MPLS LDP 功能。

（4）在 LSRA 与 LSRB、LSRC 与 LSRD 之间配置 DoD 标签发布方式（LSRB 与 LSRC 之间采用缺省的 DU 标签发布方式）。

（5）在 LSRA 和 LSRD 上配置 LDP 跨域扩展，使 LDP 按照最长匹配方式查找路由用来建立 LDP LSP。

（6）在 LSRA 和 LSRD 上配置 LDP 远端会话和 LDP 自动触发 DoD 请求功能，以尽可能地节省网络资源，减少不必要的 LSP 和 MPLS 表项。

2. 具体配置步骤

（1）配置各 LSR 上的各接口 IP 地址。

LSRA 上的配置。

```
<Huawei> system-view
[Huawei] sysname LSRA
```

```
[LSRA] interface loopback 0
[LSRA-LoopBack0] ip address 10.10.1.1 32
[LSRA-LoopBack0] quit
[LSRA] interface gigabitethernet 1/0/0
[LSRA-GigabitEthernet1/0/0] ip address 10.1.1.1 24
[LSRA-GigabitEthernet1/0/0] quit
```

LSRB 上的配置。

```
<Huawei> system-view
[Huawei] sysname LSRB
[LSRB] interface loopback 0
[LSRB-LoopBack0] ip address 10.10.1.2 32
[LSRB-LoopBack0] quit
[LSRB] interface gigabitethernet 1/0/0
[LSRB-GigabitEthernet1/0/0] ip address 10.1.1.2 24
[LSRB-GigabitEthernet1/0/0] quit
[LSRB] interface gigabitethernet 2/0/0
[LSRB-GigabitEthernet2/0/0] ip address 10.1.2.1 24
[LSRB-GigabitEthernet2/0/0] quit
```

LSRC 上的配置。

```
<Huawei> system-view
[Huawei] sysname LSRC
[LSRC] interface loopback 0
[LSRC-LoopBack0] ip address 10.10.1.3 32
[LSRC-LoopBack0] quit
[LSRC] interface gigabitethernet 1/0/0
[LSRC-GigabitEthernet1/0/0] ip address 10.1.2.2 24
[LSRC-GigabitEthernet1/0/0] quit
[LSRC] interface gigabitethernet 2/0/0
[LSRC-GigabitEthernet2/0/0] ip address 10.1.3.1 24
[LSRC-GigabitEthernet2/0/0] quit
```

LSRD 上的配置。

```
<Huawei> system-view
[Huawei] sysname LSRD
[LSRD] interface loopback 0
[LSRD-LoopBack0] ip address 10.10.1.4 32
[LSRD-LoopBack0] quit
[LSRD] interface gigabitethernet 1/0/0
[LSRD-GigabitEthernet1/0/0] ip address 10.1.3.2 24
[LSRD-GigabitEthernet1/0/0] quit
```

（2）配置骨干设备（包括 LSRB 和 LSRC）的 IS-IS 协议的基本功能，以及边缘设备和邻居的静态路由。有关 IS-IS 路由协议的配置参见《华为路由器学习指南》。

配置 LSRB 的 IS-IS 基本功能，并引入静态路由。

```
[LSRB] isis 1
[LSRB-isis-1] network-entity 10.0000.0000.0001.00    #---加入区域 2 中，系统 ID 为 1
[LSRB-isis-1] import-route static   #---引入 LSRB 上配置的到达 LSRA 的静态路由
[LSRB-isis-1] quit
[LSRB] interface gigabitethernet 2/0/0
[LSRB-GigabitEthernet2/0/0] isis enable 1   #---在 GE2/0/0 接口上使能 IS-IS 1 进程
[LSRB-GigabitEthernet2/0/0] quit
[LSRB] interface loopback 0
[LSRB-LoopBack0] isis enable 1   #---在 Loopback0 接口上使能 IS-IS 1 进程
[LSRB-LoopBack0] quit
```

配置 LSRC 的 IS-IS 基本功能，并引入静态路由。

```
[LSRC] isis 1
[LSRC-isis-1] network-entity 10.0000.0000.0002.00
[LSRC-isis-1] import-route static
[LSRC-isis-1] quit
[LSRC] interface gigabitethernet 1/0/0
[LSRC-GigabitEthernet1/0/0] isis enable 1
[LSRC-GigabitEthernet1/0/0] quit
[LSRC] interface loopback 0
[LSRC-LoopBack0] isis enable 1
[LSRC-LoopBack0] quit
```

在 LSRA 上配置下一跳为 10.1.1.2、到达 LSRB 的缺省路由。

```
[LSRA] ip route-static 0.0.0.0 0.0.0.0 10.1.1.2
```

在 LSRB 上配置到达 LSRA 的静态路由。

```
[LSRB] ip route-static 10.10.1.1 255.255.255.255 10.1.1.1
```

在 LSRC 上配置到达 LSRD 的静态路由。

```
[LSRC] ip route-static 10.10.1.4 255.255.255.255 10.1.3.2
```

在 LSRD 上配置下一跳为 10.1.3.1、到达 LSRC 的缺省路由。

```
[LSRD] ip route-static 0.0.0.0 0.0.0.0 10.1.3.1
```

以上配置完成后，在 LSRB 和 LSRC 上执行 **display ip routing-table** 命令查看路由信息可以看到，这两个 LSR 上已存在所引入的缺省路由，但没有针对 LSRA 和 LSRD 上 Loopback0 接口对应的主机路由，这样它们就不会向其上游设备发布基于这些网段的标签映射信息。

（3）在各 LSR 上配置 LSR ID（以各自的 Loopback0 接口的 IP 地址标识），使能各节点全局和公网接口的 MPLS 和 MPLS LDP 能力。

LSRA 上的配置。

```
[LSRA] mpls lsr-id 10.10.1.1
[LSRA] mpls
[LSRA-mpls] quit
[LSRA] mpls ldp
[LSRA-mpls-ldp] quit
[LSRA] interface gigabitethernet 1/0/0
[LSRA-GigabitEthernet1/0/0] mpls
[LSRA-GigabitEthernet1/0/0] mpls ldp
[LSRA-GigabitEthernet1/0/0] quit
```

LSRB 上的配置。

```
[LSRB] mpls lsr-id 10.10.1.2
[LSRB] mpls
[LSRB-mpls] quit
[LSRB] mpls ldp
[LSRB-mpls-ldp] quit
[LSRB] interface gigabitethernet 1/0/0
[LSRB-GigabitEthernet1/0/0] mpls
[LSRB-GigabitEthernet1/0/0] mpls ldp
[LSRB-GigabitEthernet1/0/0] quit
[LSRB] interface gigabitethernet 2/0/0
[LSRB-GigabitEthernet2/0/0] mpls
[LSRB-GigabitEthernet2/0/0] mpls ldp
[LSRB-GigabitEthernet2/0/0] quit
```

LSRC 上的配置。

```
[LSRC] mpls lsr-id 10.10.1.3
[LSRC] mpls
[LSRC-mpls] quit
[LSRC] mpls ldp
[LSRC-mpls-ldp] quit
[LSRC] interface gigabitethernet 1/0/0
[LSRC-GigabitEthernet1/0/0] mpls
[LSRC-GigabitEthernet1/0/0] mpls ldp
[LSRC-GigabitEthernet1/0/0] quit
[LSRC] interface gigabitethernet 2/0/0
[LSRC-GigabitEthernet2/0/0] mpls
[LSRC-GigabitEthernet2/0/0] mpls ldp
[LSRC-GigabitEthernet2/0/0] quit
```

LSRD 上的配置。

```
[LSRD] mpls lsr-id 10.10.1.4
[LSRD] mpls
[LSRD-mpls] quit
[LSRD] mpls ldp
[LSRD-mpls-ldp] quit
[LSRD] interface gigabitethernet 2/0/0
[LSRD-GigabitEthernet2/0/0] mpls
[LSRD-GigabitEthernet2/0/0] mpls ldp
[LSRD-GigabitEthernet2/0/0] quit
```

（4）同步修改 LSRA、LSRD 与邻居设备之间接口上的标签发布方式为 DoD 方式（缺省为 DU 方式），使 LSRB、LSRC 仅当 LSRA、LSRD 向它们发送标签请求消息时才给它们发送标签映射消息，分配对应的标签，建立对应的 LSP。

LSRA 上的配置。

```
[LSRA] interface gigabitethernet 1/0/0
[LSRA-GigabitEthernet1/0/0] mpls ldp advertisement dod
[LSRA-GigabitEthernet1/0/0] quit
```

LSRB 上的配置。

```
[LSRB] interface gigabitethernet 1/0/0
[LSRB-GigabitEthernet1/0/0] mpls ldp advertisement dod
[LSRB-GigabitEthernet1/0/0] quit
```

LSRC 上的配置。

```
[LSRC] interface gigabitethernet 2/0/0
[LSRC-GigabitEthernet2/0/0] mpls ldp advertisement dod
[LSRC-GigabitEthernet2/0/0] quit
```

LSRD 上的配置。

```
[LSRD] interface gigabitethernet 2/0/0
[LSRD-GigabitEthernet2/0/0] mpls ldp advertisement dod
[LSRD-GigabitEthernet2/0/0] quit
```

（5）在 LSRA、LSRD 上配置 LDP 跨域扩展功能，使能 LDP 按照最长匹配方式查找路由建立 LSP，因为 LSRA 和 LSRD 之间跨域了，不能直接学习对方的精确路由。

LSRA 上的配置。

```
[LSRA] mpls ldp
[LSRA-mpls-ldp] longest-match
[LSRA-mpls-ldp] quit
```

LSRD 上的配置。

```
[LSRD] mpls ldp
[LSRD-mpls-ldp] longest-match
[LSRD-mpls-ldp] quit
```

（6）在 LSRA 和 LSRD 上配置 LDP 远端会话，并使能 LDP 自动触发 DoD 请求功能，以请求获取远端对等体 LSR-ID 所代表的 FEC 的标签映射消息，在它们之间建立 LSP。

LSRA 上的配置

```
[LSRA] mpls ldp remote-peer lsrd   #---创建一个远端对等体
[LSRA-mpls-ldp-remote-lsrd] remote-ip 10.10.1.4    #---指定远端对等体 IP 地址为 LSRD
[LSRA-mpls-ldp-remote-lsrd] remote-ip auto-dod-request   #---使能自动触发 DoD 请求功能
[LSRA-mpls-ldp-remote-lsrd] quit
```

LSRD 上的配置

```
[LSRD] mpls ldp remote-peer lsra
[LSRD-mpls-ldp-remote-lsra] remote-ip 10.10.1.1
[LSRD-mpls-ldp-remote-lsra] remote-ip auto-dod-request
[LSRD-mpls-ldp-remote-lsra] quit
```

3. 配置结果验证

上述配置全部完成后，可在 LSRA 上执行 **display ip routing-table** 10.10.1.4 命令查看路由可以发现，路由表中没有到 10.10.1.4 的精确路由，但却可以与一条缺省路由匹配（缺省路由可与任何目标 IP 地址匹配）。

```
[LSRA] display ip routing-table 10.10.1.4
Route Flags: R - relay, D - download to fib
------------------------------------------------------------------------------
Routing Table : Public
Summary Count : 1
Destination/Mask    Proto   Pre  Cost      Flags NextHop         Interface

        0.0.0.0/0   Static 60    0           RD   10.1.1.2        GigabitEthernet1/0/0
```

在 LSRA 上执行 **display mpls ldp lsp** 命令，查看已经建立的 LSP。

可以看出，LSRA 已经建立了到达 LSRD（10.10.1.4）的 LSP（参见输出信息中的粗体字部分）。由此可见，LSRA 已经自动向 LSRB 请求了到 10.10.1.4 的标签映射消息，并且在 LSRA 与 LSRD 之间已建立了 LDP 会话。

```
[LSRA] display mpls ldp lsp

 LDP LSP Information
 ------------------------------------------------------------------------------
 DestAddress/Mask   In/OutLabel    UpstreamPeer    NextHop         OutInterface
 ------------------------------------------------------------------------------
 10.10.1.1/32       3/NULL         10.10.1.4       127.0.0.1       InLoop0
 10.10.1.4/32       NULL/1026      -               10.1.1.2        GE1/0/0
 ------------------------------------------------------------------------------
 TOTAL: 1 Normal LSP(s) Found.
 TOTAL: 0 Liberal LSP(s) Found.
 TOTAL: 0 Frr LSP(s) Found.
 A '*' before an LSP means the LSP is not established
 A '*' before a Label means the USCB or DSCB is stale
 A '*' before a UpstreamPeer means the session is stale
 A '*' before a DS means the session is stale
 A '*' before a NextHop means the LSP is FRR LSP
```

在 LSRA 或 LSRD 上执行 **display tunnel-info all** 命令可以查看 LSRA 与 LSRD

之间的隧道是否建立成功。

从显示信息可以看到，LSRA 到 LSRD 的隧道已经建立。

```
[LSRA] display tunnel-info all
 * -> Allocated VC Token
Tunnel ID          Type              Destination          Token
----------------------------------------------------------------------
0x10000001         lsp               10.10.1.4            0
```

3.3.6 配置 LDP 标签策略

一般情况下，LSR 会自动向其上游和下游 LDP 对等体分配标签，这样做的好处是可以在网络拓扑结构发生变化时提高 LDP LSP 的收敛速度。但是接收所有的标签映射消息，或者向所有对等体发送标签映射消息会导致大量 LSP 的建立，而且通常情况下，只有向上游分配的标签是有用的，这样向下游分配的标签、建立的 LSP 大多数情况下是无用的，造成资源的浪费。为了减少 LSP 的数量，节省内存，可采取如下配置策略。

（1）配置 LDP 标签过滤机制

配置 LDP Inbound 策略或者 Outbound 策略，限制标签映射消息的接收和发送。

（2）配置 LDP 水平分割策略

水平分割策略可使 LSR 只向其上游（根据具体的 FEC 路由方向）LDP 对等体分配标签，以减少本地设备建立 LSP 的数量。

下面分别介绍以上三种方案的具体配置方法。

1. 配置 LDP Inbound 策略

配置 LDP Inbound 策略，对来自对等体的标签映射消息进行过滤，仅接收允许的标签映射消息，具体的配置步骤见表 3-7。

表 3-7　配置 LDP Inbound 策略的步骤

步骤	命令	说明
1	**system-view** 例如：<Huawei> **system-view**	进入系统视图
2	**mpls ldp** 例如：[Huawei] **mpls ldp**	进入 MPLS LDP 视图
3	**inbound peer** { *peer-id* | **all** } **fec** { **none** | **host** | **ip-prefix** *prefix-name* } 例如：[Huawei-mpls-ldp] **inbound peer all fec host**	配置给指定的对等体针对指定的 IGP 路由应用 Inbound 策略，使本端仅接收来自指定对等体（或对等体组）发来的针对指定 FEC 的标签映射消息。命令中的参数选项说明如下。 • *peer-id*：多选一参数，指定的对等体 ID。缺省情况下，该参数由 **mpls lsr-id** 命令定义。如果配置了 **lsr-id** 命令，则该参数由 **lsr-id** 命令定义，点分十进制格式。 • **all**：多选一选项，所有 LDP 对等体。 • **none**：多选一选项，策略过滤掉所选定对等体发来的基于所有 FEC 的标签映射消息，即不接收指定的对等体发来的针对所有 IGP 路由的标签映射消息。 • **host**：多选一选项，策略只允许主机路由的 FEC 通过，即针对指定的对等体，仅接收由其发送的主机路由的标签映射消息。

（续表）

步骤	命令	说明
3	**inbound peer** { *peer-id* | **all** } **fec** { **none** | **host** | **ip-prefix** *prefix-name* } 例如：[Huawei-mpls-ldp] **inbound peer all fec host**	• **ip-prefix** *prefix-name*：多选一选项，策略只允许 IP 地址前缀列表指定的 FEC 通过，必须是已通过 **ip ip-prefix** *ip-prefix-name* [**index** *index-number*] { **permit** | **deny** } *ipv4-address mask-length* [**match-network**] [**greater-equal** *greater-equal-value*] [**less-equal** *less-equal-value*]命令创建的 IP 地址前缀列表。 缺省情况下，没有配置给指定的对等体针对指定的 IGP 路由应用 Inbound 策略，可用 **undo inbound peer** { *peer-id* | **peer-group** *peer-group-name* | **all** } **fec** 命令恢复缺省配置

注意 多个 Inbound 策略共存的情况下，针对某一个对等体，实际生效的 Inbound 策略以第一次的配置为准。例如先后进行了如下配置：

```
inbound peer 2.2.2.2 fec host
inbound peer peer-group group1 fec none
```

其中 group1 中包含 *peer-id* 为 2.2.2.2 的对等体，则对于 2.2.2.2 的对等体，实际生效的 Inbound 策略是：**inbound peer** 2.2.2.2 **fec host**。

如果先后配置的两条 Inbound 策略，其关键字 **peer** 部分的配置完全一样，则新的配置会覆盖旧的配置，即新的配置生效。例如先后进行了如下配置，则对于 2.2.2.2 的对等体实际生效的是：**inbound peer 2.2.2.2 fec none**。

```
inbound peer 2.2.2.2 fec host
inbound peer 2.2.2.2 fec none
```

2. 配置 LDP Outbound 策略

配置 LDP Outbound 策略，可以对本地设备向指定对等体（或对等体组）发送的标签映射消息进行过滤，仅发送允许的标签映射消息。在 Outbound 策略中不仅可以限制向指定对等体发送 IGP 路由的标签映射消息，还可限制向指定对等体发送 BGP 路由的标签映射消息，具体的配置步骤见表 3-8。

表 3-8　　配置 LDP outbound 策略的步骤

步骤	命令	说明
1	**system-view** 例如：<Huawei> **system-view**	进入系统视图
2	**mpls ldp** 例如：[Huawei] **mpls ldp**	进入 MPLS LDP 视图
3	**outbound peer** { *peer-id* | **all** } **fec** { **none** | **host** | **ip-prefix** *prefix-name* } 例如：[Huawei-mpls-ldp] **inbound peer all fec host**	（可选）配置给指定的对等体针对指定的 IGP 路由应用 Outbound 策略。命令中的参数选项说明参见表 3-7 中的第 3 步，不同的只是此处限制的是从本地向指定对等体发送的 IGP 路由标签映射消息，而不是限制来自指定对等体的 IGP 路由标签映射消息。 缺省情况下，没有配置给指定的对等体针对指定的 IGP 路由应用 Outbound 策略，可用 **undo outbound peer** { *peer-id* | **peer-group** *peer-group-name* | **all** } **fec** 命令恢复缺省配置

（续表）

步骤	命令	说明
4	**outbound peer** { *peer-id*| **all** } **bgp-label-route** { **none** | **ip-prefix** *prefix-name* } 例如：[Huawei-mpls-ldp] **outbound peer all bgp-label-route ip-prefix** prefix1	（可选）配置给指定的对等体针对指定的 BGP 标签路由应用 Outbound 策略。命令中的参数和选项说明如下。 • *peer-id*：多选一参数，指定对等体 ID。缺省情况下，该参数由 **mpls lsr-id** 命令定义。如果配置了 **lsr-id** 命令，则该参数由 **lsr-id** 命令定义。 • **all**：多选一选项，所有 LDP 对等体。 • **none**：二选一选项，策略过滤掉所有 FEC，即不给指定的对等体发送 BGP 路由的标签映射消息。 • **ip-prefix** *prefix-name*：二选一参数，策略只允许 IP 地址前缀列表指定的 FEC 通过，即给指定的对等体发送 IP 地址前缀列表规定 BGP 路由的标签映射消息。 缺省情况下，没有根据指定的 BGP 标签路由配置 Outbound 策略，不会给指定的对等体发送标签映射消息，可用 **undo outbound peer** { *peer-id* | **peer-group** *peer-group-name* | **all** } **bgp-label-route** 命令恢复缺省配置

注意　多个 Outbound 策略共存的情况下，针对某一个对等体，实际生效的 Outbound 策略以第一次的配置为准。例如先后进行了如下配置：

```
outbound peer 2.2.2.2 bgp-label-route ip-prefix prefix1
outbound peer peer-group group1 bgp-label-route none
```

其中 group1 中包含 *peer-id* 为 2.2.2.2 的对等体，对于该对等体，实际生效的 Outbound 策略是：**outbound peer 2.2.2.2 bgp-label-route ip-prefix prefix1**。

如果先后配置的两条 Outbound 策略，其关键字 **peer** 部分的配置完全一样，则新的配置会覆盖旧的配置。例如先后进行了如下配置，则对于 2.2.2.2 对等体来说，实际生效的 Outbound 策略是：**outbound peer** 2.2.2.2 **bgp-label-route none**。

```
outbound peer 2.2.2.2 bgp-label-route ip-prefix prefix1
outbound peer 2.2.2.2 bgp-label-route none
```

3. 配置 LDP 水平分割

LDP 对等体配置水平分割策略可使 LSR 只向其上游（根据对应 FEC LSP 方向）LDP 对等体分配标签。配置方法很简单，只需在 MPLS LDP 视图下通过 **outbound peer** { *peer-id* | **all** } **split-horizon** 命令配置即可。命令中的参数和选项说明如下。

■ *peer-id*：二选一参数，指定不给指定 LDR ID 的下游 LDP 对等体分配标签。

■ **all**：二选一选项，指定不给所有下游 LDP 对等体分配标签。

缺省情况下，没有为 LDP 对等体配置水平分割策略，即 LSR 会向其上游和下游 LDP 对等体都分配标签，可用 **undo outbound peer** { *peer-id* | **all** } **split-horizon** 命令恢复缺省配置。

注意　针对所有 peer 的 LDP 水平分割策略比针对某个 peer 的 LDP 水平分割策略优先级高。例如先配置 **outbound peer all split-horizon**，然后再配置 **outbound peer** 2.2.2.2 **split-horizon**，则单个 peer 的 LDP 水平分割策略不生效。

3.3.7 LDP Inbound 策略配置示例

如图 3-13 所示网络中，部署了 MPLS LDP 业务，LSRD 是接入设备，性能较低。如果不对 LSRD 收到的标签进行控制，则会建立大量的 LSP，从而消耗大量内存，LSRD 无法承受。要求有效地减少 LSP 的数量，从而节省 LSRD 内存，减少资源的浪费。

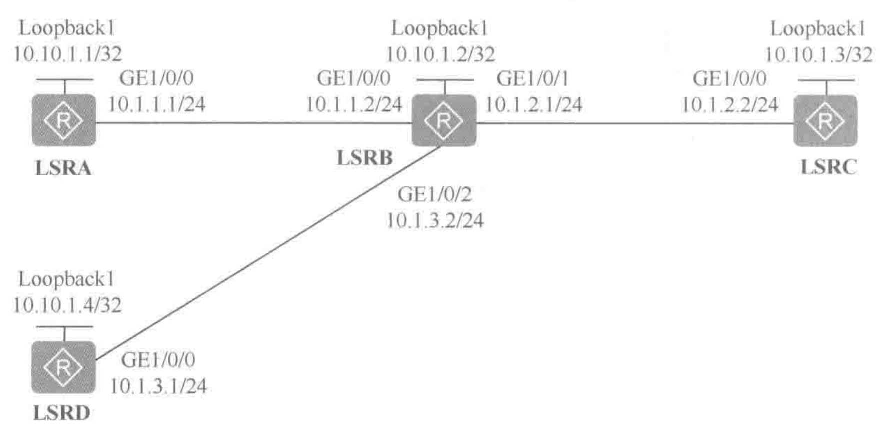


图 3-13 LDP Inbound 策略配置示例的拓扑结构

1. 基本配置思路分析

本示例是要限制 LSRD 建立 LSP 的数量，也就是要限制 LSRD 从对等体接收的标签映射消息。这时可通过配置 LDP Inbound 策略实现此需求。

LDP Inbound 策略是 LDP 可选基本功能，在配置此任务前要完成 LDP 必选基本功能的配置，并且要确保各 LSR 的路由畅通。由此可得出本示例的基本配置路由如下。

（1）在各 LSR 上配置各接口（包括 Loopback 接口）的 IP 地址。

（2）在各 LSR 上配置 OSPF 路由（包括 Loopback 接口主机路由）。

（3）在各 LSR 上配置 LSR ID，全局使能 MPLS 和 MPLS LDP 功能，以及在各 LSR 互连的公网接口上使能 MPLS 和 MPLS LDP 功能。

（4）在 LSRD 上配置 LDP Inbound 策略，只接收由 LSRB 发送的、到达 LSRC 的标签映射消息。

2. 具体配置步骤

（1）配置各接口的 IP 地址。

\# LSRA 上的配置。

```
<Huawei> system-view
[Huawei] sysname LSRA
[LSRA] interface loopback 1
[LSRA-LoopBack1] ip address 10.10.1.1 32
[LSRA-LoopBack1] quit
[LSRA] interface gigabitethernet 1/0/0
[LSRA-GigabitEthernet1/0/0] ip address 10.1.1.1 24
[LSRA-GigabitEthernet1/0/0] quit
```

\# LSRB 上的配置。

```
<Huawei> system-view
[Huawei] sysname LSRB
[LSRB] interface loopback 1
```

```
[LSRB-LoopBack1] ip address 10.10.1.2 32
[LSRB-LoopBack1] quit
[LSRB] interface gigabitethernet 1/0/0
[LSRB-GigabitEthernet1/0/0] ip address 10.1.1.2 24
[LSRB-GigabitEthernet1/0/0] quit
[LSRB] interface gigabitethernet 1/0/1
[LSRB-GigabitEthernet1/0/1] ip address 10.1.2.1 24
[LSRB-GigabitEthernet1/0/1] quit
[LSRB] interface gigabitethernet 1/0/2
[LSRB-GigabitEthernet1/0/2] ip address 10.1.3.2 24
[LSRB-GigabitEthernet1/0/2] quit
```

LSRC 上的配置。

```
<Huawei> system-view
[Huawei] sysname LSRC
[LSRC] interface loopback 1
[LSRC-LoopBack1] ip address 10.10.1.3 32
[LSRC-LoopBack1] quit
[LSRC] interface gigabitethernet 1/0/0
[LSRC-GigabitEthernet1/0/0] ip address 10.1.2.2 24
[LSRC-GigabitEthernet1/0/0] quit
```

LSRD 上的配置。

```
<Huawei> system-view
[Huawei] sysname LSRD
[LSRD] interface loopback 1
[LSRD-LoopBack1] ip address 10.10.1.3 32
[LSRD-LoopBack1] quit
[LSRD] interface gigabitethernet 1/0/0
[LSRD-GigabitEthernet1/0/0] ip address 10.1.3.1 24
[LSRD-GigabitEthernet1/0/0] quit
```

（2）配置 OSPF 协议发布各节点接口所连网段和 LSR ID 的主机路由。各 LSR 均同在 OSPF 路由进程 1、区域 0 中。

LSRA 上的配置。

```
[LSRA] ospf 1
[LSRA-ospf-1] area 0
[LSRA-ospf-1-area-0.0.0.0] network 10.10.1.1 0.0.0.0
[LSRA-ospf-1-area-0.0.0.0] network 10.1.1.0 0.0.0.255
[LSRA-ospf-1-area-0.0.0.0] quit
[LSRA-ospf-1] quit
```

LSRB 上的配置。

```
[LSRB] ospf 1
[LSRB-ospf-1] area 0
[LSRB-ospf-1-area-0.0.0.0] network 10.10.1.2 0.0.0.0
[LSRB-ospf-1-area-0.0.0.0] network 10.1.1.0 0.0.0.255
[LSRB-ospf-1-area-0.0.0.0] network 10.1.2.0 0.0.0.255
[LSRB-ospf-1-area-0.0.0.0] network 10.1.3.0 0.0.0.255
[LSRB-ospf-1-area-0.0.0.0] quit
[LSRB-ospf-1] quit
```

LSRC 上的配置。

```
[LSRC] ospf 1
[LSRC-ospf-1] area 0
[LSRC-ospf-1-area-0.0.0.0] network 10.10.1.3 0.0.0.0
[LSRC-ospf-1-area-0.0.0.0] network 10.1.2.0 0.0.0.255
```

```
[LSRC-ospf-1-area-0.0.0.0] quit
[LSRC-ospf-1] quit
```

LSRD 上的配置。

```
[LSRD] ospf 1
[LSRD-ospf-1] area 0
[LSRD-ospf-1-area-0.0.0.0] network 10.10.1.4 0.0.0.0
[LSRD-ospf-1-area-0.0.0.0] network 10.1.3.0 0.0.0.255
[LSRD-ospf-1-area-0.0.0.0] quit
[LSRD-ospf-1] quit
```

（3）配置各 LSR 的 LSR ID，使能全局和公网接口（不包括 Loopback 接口）的 MPLS 和 MPLS LDP 能力。

LSRA 上的配置。

```
[LSRA] mpls lsr-id 10.10.1.1
[LSRA] mpls
[LSRA-mpls] quit
[LSRA] mpls ldp
[LSRA-mpls-ldp] quit
[LSRA] interface gigabitethernet 1/0/0
[LSRA-GigabitEthernet1/0/0] mpls
[LSRA-GigabitEthernet1/0/0] mpls ldp
[LSRA-GigabitEthernet1/0/0] quit
```

LSRB 上的配置。

```
[LSRB] mpls lsr-id 10.10.1.2
[LSRB] mpls
[LSRB-mpls] quit
[LSRB] mpls ldp
[LSRB-mpls-ldp] quit
[LSRB] interface gigabitethernet 1/0/0
[LSRB-GigabitEthernet1/0/0] mpls
[LSRB-GigabitEthernet1/0/0] mpls ldp
[LSRB-GigabitEthernet1/0/0] quit
[LSRB] interface gigabitethernet 1/0/1
[LSRB-GigabitEthernet1/0/1] mpls
[LSRB-GigabitEthernet1/0/1] mpls ldp
[LSRB-GigabitEthernet1/0/1] quit
[LSRB] interface gigabitethernet 1/0/2
[LSRB-GigabitEthernet1/0/2] mpls
[LSRB-GigabitEthernet1/0/2] mpls ldp
[LSRB-GigabitEthernet1/0/2] quit
```

LSRC 上的配置。

```
[LSRC] mpls lsr-id 10.10.1.3
[LSRC] mpls
[LSRC-mpls] quit
[LSRC] mpls ldp
[LSRC-mpls-ldp] quit
[LSRC] interface gigabitethernet 1/0/0
[LSRC-GigabitEthernet1/0/0] mpls
[LSRC-GigabitEthernet1/0/0] mpls ldp
[LSRC-GigabitEthernet1/0/0] quit
```

LSRD 上的配置。

```
[LSRD] mpls lsr-id 10.10.1.4
[LSRD] mpls
```

```
[LSRD-mpls] quit
[LSRD] mpls ldp
[LSRD-mpls-ldp] quit
[LSRD] interface gigabitethernet 1/0/0
[LSRD-GigabitEthernet1/0/0] mpls
[LSRD-GigabitEthernet1/0/0] mpls ldp
[LSRD-GigabitEthernet1/0/0] quit
```

完成以上配置后，在 LSRD 上执行 **display mpls lsp** 命令，查看已经建立的 LSP 会发现 LSRD 上建立了到 LSRA、LSRB、LSRC 的 LSP，而且各有两条 LSP（参见输出信息中的粗体字部分）。

```
[LSRD] display mpls lsp
-------------------------------------------------------------------------------
                 LSP Information: LDP LSP
-------------------------------------------------------------------------------
FEC                In/Out Label    In/Out IF                     Vrf Name
10.10.1.1/32       NULL/1024        -/GE1/0/0
10.10.1.1/32       1024/1024       -/GE1/0/0
10.10.1.2/32       NULL/3           -/GE1/0/0
10.10.1.2/32       1025/3          -/GE1/0/0
10.10.1.3/32       NULL/1025        -/GE1/0/0
10.10.1.3/32       1026/1025       -/GE1/0/0
10.10.1.4/32       3/NULL           -/-
```

在这些基于同一 FEC 的两条 LSP 中，其中只带有出标签的那条 LSP 是作为从本地到达指定目的地址的 LSP，而另一条同时包括入标签和出标签的 LSP 则是作为指导其上游设备到达目的地址的 LSP。

（4）配置 LDP Inbound 策略。

在 LSRD 上配置 IP 地址前缀列表，只允许到 LSRC 的路由通过。

```
[LSRD] ip ip-prefix prefix1 permit 10.10.1.3 32
```

在 LSRD 上配置 Inbound 策略，使其只接收由 LSRB 发送的到 LSRC 10.10.1.3/32 的标签映射消息。

```
[LSRD] mpls ldp
[LSRD-mpls-ldp] inbound peer 10.10.1.2 fec ip-prefix prefix1
[LSRD-mpls-ldp] quit
```

配置完成后，再在 LSRD 上执行 **display mpls lsp** 命令，此时就可以看到除了为本地 10.10.1.4/32 建立的 LSP 外，就仅建立了到 LSRC 10.10.1.3/32 的两条 LSP（参见输出信息粗体字部分），原来到达 LSRA 和 LSRB 的 4 条 LSP 全部被过滤了。

```
[LSRD] display mpls lsp
-------------------------------------------------------------------------------
                 LSP Information: LDP LSP
-------------------------------------------------------------------------------
FEC                In/Out Label    In/Out IF                     Vrf Name
10.10.1.3/32       NULL/1025        -/GE1/0/0
10.10.1.3/32       1026/1025       -/GE1/0/0
10.10.1.4/32       3/NULL           -/-
```

这样一来，LSRD 也只能通过 MPLS 标签交换方式访问 LSRC 的 10.10.1.3/32 主机，而不能通过 MPLS 标签交换方式访问 LSRA 的 10.10.1.1/32 和 LSRB 的 10.1.1.2/32，但仍可以通过路由方式访问。

3.3.8 LDP Outbound 策略配置示例

本示例的网络拓扑结构仍参见 3.3.7 节的图 3-13，在网络中部署了 MPLS LDP 业务，LSRD 是接入设备，性能较低。如果不对 LSRD 收到的标签进行控制，则会建立大量的 LSP，消耗大量内存，LSRD 无法承受。要求有效地减少 LSP 的数量，从而节省 LSRD 内存，减少资源的浪费。

1. 基本配置思路分析

本示例与 3.3.7 节介绍的配置示例的总体目标是一样的，就是想让 LSRD 上建立的 LSP 少些，但本示例要求采用的是 LDP Outbound 策略，所以配置的对象与 3.3.7 节介绍的配置示例不一样。3.3.7 节示例是在 LSRD 上配置 LDP Inbound 策略，用以限制 LSRD 所接收的标签映射消息，本示例要在与 LSRD 直接相连的 LSRB 上配置 LDP Outbound 策略，使 LSRB 仅向 LSRD 发送 LSRD 所需的标签映射消息。

本示例的总体配置思路与 3.3.7 节配置示例差不多，只是最后的 LDP 过滤策略配置不一样，具体如下。

（1）在各 LSR 上配置各接口（包括 Loopback 接口）的 IP 地址。

（2）在各 LSR 上配置 OSPF 路由（包括 Loopback 接口主机路由）。

（3）在各 LSR 上配置 LSR ID，全局使能 MPLS 和 MPLS LDP 功能，以及在各 LSR 互连的接口上使能 MPLS 和 MPLS LDP 功能。

（4）在 LSRB 上配置 LDP Outbound 策略，使其只向 LSRD 发送到达 LSRC 10.10.1.3/32 的标签映射消息。

2. 具体配置步骤

因为本示例的拓扑结构与 3.3.7 节图 3-13 是一样的，所以上述配置思路中的第（1）～（3）项配置任务的具体配置与 3.3.7 节完全相同，所以在此仅介绍上述第（4）项配置任务的具体配置方法。

在完成第（1）～（3）配置任务后，在 LSRD 上执行 **display mpls lsp** 命令，可以查看到已经建立了到达各 LSR 的 LSP，具体如下所示。

```
[LSRD] display mpls lsp
-------------------------------------------------------------------------------
                 LSP Information: LDP LSP
-------------------------------------------------------------------------------
FEC              In/Out Label    In/Out IF                      Vrf Name
10.10.1.1/32     NULL/1024        -/GE1/0/0
10.10.1.1/32     1024/1024       -/GE1/0/0
10.10.1.2/32     NULL/3           -/GE1/0/0
10.10.1.2/32     1025/3          -/GE1/0/0
10.10.1.3/32     NULL/1025        -/GE1/0/0
10.10.1.3/32     1026/1025       -/GE1/0/0
10.10.1.4/32     3/NULL           -/-
```

（4）配置 LDP Outbound 策略。

在 LSRB 上配置 IP 地址前缀列表，只允许到 LSRC 的路由通过。

```
[LSRB] ip ip-prefix prefix1 permit 10.10.1.3 32
```

在 LSRB 上配置 Outbound 策略，使其只给 LSRD 发送到 LSRC 的标签映射消息。

```
[LSRB] mpls ldp
```

```
[LSRB-mpls-ldp] outbound peer 10.10.1.4 fec ip-prefix prefix1
[LSRB-mpls-ldp] quit
```

配置好后，再在LSRD上执行 **display mpls lsp** 命令，此时可以看到除了本地直连的10.10.1.4/32 LSP外，也只建立了到达LSRC的两条LSP（原来到达LSRA和LSRB的4条LSP没有了，参见输出信息粗体字部分），实现了与3.3.7节配置示例一样的效果。

```
[LSRD] display mpls lsp
-------------------------------------------------------------------------------
                  LSP Information: LDP LSP
-------------------------------------------------------------------------------
FEC              In/Out Label   In/Out IF                    Vrf Name
10.10.1.3/32     NULL/1025      -/GE1/0/0
10.10.1.3/32     1026/1025      -/GE1/0/0
10.10.1.4/32     3/NULL         -/-
```

3.3.9 配置LDP LSP建立的触发策略

缺省情况下，使能MPLS LDP后，各设备上的32位主机路由将自动建立LSP。如果不通过策略控制，将建立有大量的LSP，而其中又包括许多当前无用甚至建立不成功的LSP，导致资源浪费。

为了节省设备资源，除了可以通过采用在3.3.6节介绍的LDP标签策略过滤设备所接收或发送的标签映射消息来实现对LSP建立的控制外，还可采用本节介绍的LSP建立触发策略进行控制。

在不同节点上可配置的LDP LSP建立触发策略不一样。

- 在Ingress和Egress上配置lsp-trigger策略（具体配置步骤见表3-9），使仅符合条件的路由触发LSP的建立。
- 在Transit上配置propagate mapping策略，仅允许符合过滤条件的路由的标签映射消息向上游发送，可以有效减少上游LSP的数量，节约网络资源。但propagate mapping策略**也仅可限制非本地直连路由的标签映射消息向上游发送**，对本地直连的路由不起作用。具体的配置步骤见表3-10。

通常情况下，建议配置lsp-trigger策略。若由于某种特殊原因在Ingress和Egress上不能配置策略，则配置propagate mapping策略。

表3-9　　在Ingress和Egress上配置lsp-trigger策略的步骤

步骤	命令	说明
1	**system-view** 例如：<Huawei> **system-view**	进入系统视图
2	**mpls** 例如：[Huawei] **mpls**	进入MPLS视图
3	**lsp-trigger** { **all** | **host** | **ip-prefix** *ip-prefix-name* | **none** } 例如：[Huawei-mpls] **lsp-trigger ip-prefix** ipprefix1	（二选一）配置触发静态路由和IGP路由建立LSP的策略。命令中的参数和选项说明如下。 • **all**：多选一选项，指定在MPLS域内的静态和IGP路由都将触发建立LSP，不推荐采用。 • **host**：多选一选项，指定仅MPLS域内的32位掩码的主机IP路由（不包括接口的32位地址掩码的主机IP路由）触发建立LSP，**这是缺省选项**

（续表）

步骤	命令	说明
3	**lsp-trigger** { **all** | **host** | **ip-prefix** *ip-prefix-name* | **none** } 例如：[Huawei-mpls] **lsp-trigger ip-prefix** ipprefix1	• **ip-prefix** *ip-prefix-name*：多选一参数，指定根据 IP 地址前缀列表触发建立 LSP。最终结果是：凡是不在 IP 地址前缀列表许可范围中的路由，以及所有以该节点为 Ingress 的其他路由都将被禁止建立 LSP。 • **none**：多选一选项，不触发建立 LSP，**但不能限制本地直连路由的 LSP 建立**。 【注意】本命令只对公网的 Ingress LSP 和 Egress LSP，以及私网的 IGP 路由的 Ingress LSP 和 Egress LSP 有效。**配置触发建立 LSP 的策略为 host 时**（这是缺省配置），在不同的节点执行命令，配置效果也不同：在 Ingress 节点执行该命令时，触发 MPLS 域所有的 32 位掩码路由建立 LDP LSP；在 Egress 节点执行该命令时，触发本地 32 位掩码路由建立 LDP LSP。 缺省情况下，触发策略为 host，即 32 位地址掩码的主机 IP 路由（不包括接口的 32 位地址掩码的主机 IP 路由）触发建立 LSP，可用 **undo lsp-trigger** 命令恢复缺省设置
	lsp-trigger bgp-label-route [**ip-prefix** *ip-prefix-name*] 例如：[Huawei-mpls] **lsp-trigger bgp-label-route**	（二选一）配置触发带标签的公网 BGP 路由建立 LSP 的策略。可选参数 **ip-prefix** *ip-prefix-name* 允许通过指定 IP 地址前缀列表过滤的带标签的公网 BGP 路由触发 LDP 建立 LSP。有关 BGP 路由携带 MPLS 标签的配置和应用参见《华为 MPLS VPN 学习指南》一书。 缺省情况下，LDP 不为带标签的公网 BGP 路由分标签，可用 **undo lsp-trigger bgp-label-route** 命令恢复为缺省设置
4	**proxy-egress disable** 例如：[Huawei-mpls] **proxy-egress disable**	（可选）配置禁止建立代理 Egress LSP。当在第 3 步配置的 LSP 触发策略为所有静态路由和 IGP 路由项（选择 **all** 选项时）触发建立 LSP 或根据 IP 地址前缀列表（选择 **ip-prefix** 参数时）触发建立 LSP 时，会触发建立代理 Egress LSP。但这些代理 Egress LSP 很可能是无用的，会耗费系统资源。此时可以执行本命令禁止建立代理 Egress LSP。 缺省情况下，系统允许建立代理 Egress LSP，可用 **undo proxy-egress disable** 命令配置允许建立代理 Egress LSP

表 3-10　　在 Transit 上配置 propagate mapping 策略的步骤

步骤	命令	说明
1	**system-view** 例如：<Huawei> **system-view**	进入系统视图
2	**mpls ldp** 例如：[Huawei] **mpls ldp**	进入 MPLS LDP 视图
3	**propagate mapping for ip-prefix** *ip-prefix-name* 例如：[Huawei-mpls-ldp] **propagate mapping for ip-prefix** policy1	配置 LSP 建立策略。参数 *ip-prefix-name* 指定用于路由过滤的 IP 地址前缀列表，使仅发送符合该 IP 地址前缀列表的路由的标签映射消息给上游，需事先建立好对应的 IP 地址前缀列表。**但不能限制本地直连路由的标签映射消息发送给上游，启用了 LDP 功能的接口对应的网段不会生成标签映射消息**。 缺省情况下，LDP 在建立 LSP 时，不对收到的路由进行过滤，可用 **undo propagate mapping** 命令恢复为缺省配置

3.3.10　LSP 建立的 lsp-trigger 触发策略配置示例

在如图 3-14 所示的 MPLS 网络中，各 LSR 接口上使能 MPLS LDP 后，LDP LSP 将自动建立。如果网络规模比较大，则在各 LSR 上建立大量的 LSP，导致资源的浪费。现要求能够控制 LSP 建立的数量，从而减少系统资源的浪费。

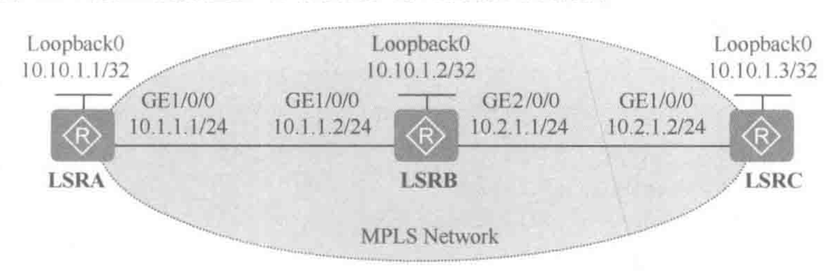


图 3-14　LSP 建立的 lsp-trigger 触发策略配置示例的拓扑结构

1. 基本配置思路分析

本示例采用配置 lsp-trigger 策略来实现，但前提也必须是先完成 MPLS LDP 的必选基本功能的配置，基本的配置思路如下。

（1）配置各 LSR 各接口（包括 Loopback 接口）的 IP 地址。

（2）配置各 LSR 的路由，实现网络互通。本示例采用 OSPF 路由协议。

（3）在各 LSR 上使有全局、公网接口的 MPLS 和 LDP 功能。

（4）在 LSRA 上配置 lsp-trigger 策略，假设只允许建立 10.10.1.3/32 的 LDP LSP。

2. 具体配置步骤

（1）配置各接口（包括 Loopback0）的 IP 地址。

LSRA 上的配置。

```
<Huawei> system-view
[Huawei] sysname LSRA
[LSRA] interface loopback 0
[LSRA-LoopBack0] ip address 10.10.1.1 32
[LSRA-LoopBack0] quit
[LSRA] interface gigabitethernet 1/0/0
[LSRA-GigabitEthernet1/0/0] ip address 10.1.1.1 24
[LSRA-GigabitEthernet1/0/0] quit
```

LSRB 上的配置。

```
<Huawei> system-view
[Huawei] sysname LSRB
[LSRB] interface loopback 0
[LSRB-LoopBack0] ip address 10.10.1.2 32
[LSRB-LoopBack0] quit
[LSRB] interface gigabitethernet 1/0/0
[LSRB-GigabitEthernet1/0/0] ip address 10.1.1.2 24
[LSRB-GigabitEthernet1/0/0] quit
[LSRB] interface gigabitethernet 2/0/0
[LSRB-GigabitEthernet2/0/0] ip address 10.2.1.1 24
[LSRB-GigabitEthernet2/0/0] quit
```

LSRC 上的配置。

```
<Huawei> system-view
[Huawei] sysname LSRC
```

```
[LSRC] interface loopback 0
[LSRC-LoopBack0] ip address 10.10.1.3 32
[LSRC-LoopBack0] quit
[LSRC] interface gigabitethernet 1/0/0
[LSRC-GigabitEthernet1/0/0] ip address 10.2.1.2 24
[LSRC-GigabitEthernet1/0/0] quit
```

（2）配置 OSPF 协议发布各节点公网接口所连网段和 LSR ID 的主机路由，加入 OSPF 路由进程 1、区域 0 中。

LSRA 上的配置。

```
[LSRA] ospf 1
[LSRA-ospf-1] area 0
[LSRA-ospf-1-area-0.0.0.0] network 10.10.1.1 0.0.0.0
[LSRA-ospf-1-area-0.0.0.0] network 10.1.1.0 0.0.0.255
[LSRA-ospf-1-area-0.0.0.0] quit
[LSRA-ospf-1] quit
```

LSRB 上的配置。

```
[LSRB] ospf 1
[LSRB-ospf-1] area 0
[LSRB-ospf-1-area-0.0.0.0] network 10.10.1.2 0.0.0.0
[LSRB-ospf-1-area-0.0.0.0] network 10.1.1.0 0.0.0.255
[LSRB-ospf-1-area-0.0.0.0] network 10.2.1.0 0.0.0.255
[LSRB-ospf-1-area-0.0.0.0] quit
[LSRB-ospf-1] quit
```

LSRC 上的配置。

```
[LSRC] ospf 1
[LSRC-ospf-1] area 0
[LSRC-ospf-1-area-0.0.0.0] network 10.10.1.3 0.0.0.0
[LSRC-ospf-1-area-0.0.0.0] network 10.2.1.0 0.0.0.255
[LSRC-ospf-1-area-0.0.0.0] quit
[LSRC-ospf-1] quit
```

以上配置完成后，在各节点上执行 **display ip routing-table** 命令，可以看到相互之间都学到了彼此的路由。

（3）配置 MPLS LDP。以各自的 Loopback0 接口 IP 地址作为它们的 MPLS LSR ID，在全局及接口上使能 MPLS 和 LDP 功能。

LSRA 上的配置。

```
[LSRA] mpls lsr-id 10.10.1.1
[LSRA] mpls
[LSRA-mpls] quit
[LSRA] mpls ldp
[LSRA-mpls-ldp] quit
[LSRA] interface gigabitethernet 1/0/0
[LSRA-GigabitEthernet1/0/0] mpls
[LSRA-GigabitEthernet1/0/0] mpls ldp
[LSRA-GigabitEthernet1/0/0] quit
```

LSRB 上的配置。

```
[LSRB] mpls lsr-id 10.10.1.2
[LSRB] mpls
[LSRB-mpls] quit
[LSRB] mpls ldp
[LSRB-mpls-ldp] quit
```

```
[LSRB] interface gigabitethernet 1/0/0
[LSRB-GigabitEthernet1/0/0] mpls
[LSRB-GigabitEthernet1/0/0] mpls ldp
[LSRB-GigabitEthernet1/0/0] quit
[LSRB] interface gigabitethernet 2/0/0
[LSRB-GigabitEthernet2/0/0] mpls
[LSRB-GigabitEthernet2/0/0] mpls ldp
[LSRB-GigabitEthernet2/0/0] quit
```

LSRC 上的配置。

```
[LSRC] mpls lsr-id 10.10.1.3
[LSRC] mpls
[LSRC-mpls] quit
[LSRC] mpls ldp
[LSRC-mpls-ldp] quit
[LSRC] interface gigabitethernet 1/0/0
[LSRC-GigabitEthernet1/0/0] mpls
[LSRC-GigabitEthernet1/0/0] mpls ldp
[LSRC-GigabitEthernet1/0/0] quit
```

以上配置完成后，各 LSR 已根据默认的 LDP LSP 触发策略，即 32 位地址掩码的主机 IP 路由都已触发建立 LDP LSP 了，这是缺省配置。在各 LSR 上执行 **display mpls ldp lsp** 命令可以看到，所有主机路由都触发建立了 LDP LSP。以下是在 LSRA 上执行该命令的输出示例。那些入标签为“Liberal”的 LSP 表示没有建立成功的 LSP。

```
[LSRA] display mpls ldp lsp

 LDP LSP Information
 -------------------------------------------------------------------------------
 DestAddress/Mask   In/OutLabel    UpstreamPeer    NextHop          OutInterface
 -------------------------------------------------------------------------------
 10.10.1.1/32       3/NULL         10.10.1.2       127.0.0.1        InLoop0
*10.10.1.1/32       Liberal/3                      DS/10.10.1.2
 10.10.1.2/32       NULL/3         -               10.1.1.2         GE1/0/0
 10.10.1.2/32       1024/3         10.10.1.2       10.1.1.2         GE1/0/0
 10.10.1.3/32       NULL/1025      -               10.1.1.2         GE1/0/0
 10.10.1.3/32       1022/1025      10.10.1.2       10.1.1.2         GE1/0/0
 -------------------------------------------------------------------------------
 TOTAL: 5 Normal LSP(s) Found.
 TOTAL: 1 Liberal LSP(s) Found.
 TOTAL: 0 Frr LSP(s) Found.
 A '*' before an LSP means the LSP is not established
 A '*' before a Label means the USCB or DSCB is stale
 A '*' before a UpstreamPeer means the session is stale
 A '*' before a DS means the session is stale
 A '*' before a NextHop means the LSP is FRR LSP
```

（4）在 LSRA 上使用 IP 前缀列表对可建立 LSP 的路由进行过滤。

在 LSRA 配置 IP 前缀列表，只允许建立 LSRC 上的 10.10.1.3/32 的 LSP。

```
[LSRA] ip ip-prefix FilterOnIngress permit 10.10.1.3 32
[LSRA] mpls
[LSRA-mpls] lsp-trigger ip-prefix FilterOnIngress
[LSRA-mpls] quit
```

3. 实验结果验证

以上配置完成后，再在各 LSR 上执行 **display mpls ldp lsp** 命令可以看到，它们上面

的 LDP LSP 的建立情况发生了很大变化，结果就是：LSRA 节点上只存在以 LSRA 为 Ingress 的关于 10.10.1.3/32 的 LDP LSP，以及其他不是以 LSRA 为 Ingress 的 LDP LSP。以下是在 LSRA 配置了 lsp-trigger 策略后的 LDP LSP 建立情况。

```
[LSRA] display mpls ldp lsp

 LDP LSP Information
 -------------------------------------------------------------------------------
 DestAddress/Mask   In/OutLabel    UpstreamPeer    NextHop         OutInterface
 -------------------------------------------------------------------------------
 10.10.1.2/32       1024/3         10.10.1.2       10.1.1.2        GE1/0/0
 10.10.1.3/32       NULL/1025      -               10.1.1.2        GE1/0/0
 10.10.1.3/32       1022/1025      10.10.1.2       10.1.1.2        GE1/0/0
 -------------------------------------------------------------------------------
 TOTAL: 3 Normal LSP(s) Found.
 TOTAL: 0 Liberal LSP(s) Found.
 TOTAL: 0 Frr LSP(s) Found.
 A '*' before an LSP means the LSP is not established
 A '*' before a Label means the USCB or DSCB is stale
 A '*' before a UpstreamPeer means the session is stale
 A '*' before a DS means the session is stale
 A '*' before a NextHop means the LSP is FRR LSP
```

从以上输出信息可以看出，在 LSRA 上原来基于 LSRA 本地 10.10.1.1/32 主机路由建立的以下两条 LSP 没有了。下面分析具体的原因。

```
DestAddress/Mask   In/OutLabel    UpstreamPeer    NextHop         OutInterface
-------------------------------------------------------------------------------
10.10.1.1/32       3/NULL         10.10.1.2       127.0.0.1       InLoop0
*10.10.1.1/32      Liberal/3                      DS/10.10.1.2
```

上面第一条 LSP 是一条本地 LSP，因为其下一跳为 127.0.0.1（代表本地），出接口为 Loopback 接口，不能算是 LDP LSP，所以被禁止建立了。上面第二条本身就是一条没有建立成功的 LSP（入标签为 Liberal），所以也被禁止了。

原来在 LSRA 上建立的以下另外 4 条 LSP 中，只有第一条没有了，其他三条均仍存在，下面进行分析。

```
DestAddress/Mask   In/OutLabel    UpstreamPeer    NextHop         OutInterface
-------------------------------------------------------------------------------
 10.10.1.2/32       NULL/3         -               10.1.1.2        GE1/0/0
 10.10.1.2/32       1024/3         10.10.1.2       10.1.1.2        GE1/0/0
 10.10.1.3/32       NULL/1025      -               10.1.1.2        GE1/0/0
 10.10.1.3/32       1022/1025      10.10.1.2       10.1.1.2        GE1/0/0
```

上面第一条因为只有出标签，显然是以 LSRA 作为 Ingress 的，但它不在 lsp-trigger 策略配置的 IP 地址前缀表列的许可范围中，所以被禁止了。上面第二条，因为同时有入标签和出标签，LSRA 是作为 Transit 的，它是不会被禁止的，所以仍然可以建立。上面第三条也只有出标签，LSRA 也是作为 Ingress，但它是在 lsp-trigger 策略配置的 IP 地址前缀列表的许可范围中，所以允许被建立。上面第四条，同时带有入标签和出标签，LSRA 是作为 Transit 的，它是不会被禁止的，所以仍然可以建立。

在 LSRA 上配置好前面的 lsp-trigger 策略后，影响的可能不仅是 LSRA 自己，对 LSRB 和 LSRC 的 LSP 建立也可会产生一些影响。以下是在 LSRA 上配置以上 lsp-trigger 策略前，在 LSRB 上执行 **display mpls ldp lsp** 命令后的输出信息。

```
<LSRB>display mpls ldp lsp

 LDP LSP Information
 -------------------------------------------------------------------------------
 DestAddress/Mask   In/OutLabel    UpstreamPeer    NextHop         OutInterface
 -------------------------------------------------------------------------------

 10.10.1.1/32       NULL/3         -               10.1.1.1        GE1/0/0
 10.10.1.1/32       1029/3         10.10.1.1       10.1.1.1        GE1/0/0
 10.10.1.1/32       1029/3         10.10.1.3       10.1.1.1        GE1/0/0
*10.10.1.1/32       Liberal/1029                   DS/10.10.1.3
 10.10.1.2/32       3/NULL         10.10.1.1       127.0.0.1       InLoop0
 10.10.1.2/32       3/NULL         10.10.1.3       127.0.0.1       InLoop0
*10.10.1.2/32       Liberal/1024                   DS/10.10.1.1
*10.10.1.2/32       Liberal/1025                   DS/10.10.1.3
 10.10.1.3/32       NULL/3         -               10.2.1.2        GE2/0/0
 10.10.1.3/32       1026/3         10.10.1.1       10.2.1.2        GE2/0/0
 10.10.1.3/32       1026/3         10.10.1.3       10.2.1.2        GE2/0/0
*10.10.1.3/32       Liberal/1025                   DS/10.10.1.1
 -------------------------------------------------------------------------------
 TOTAL: 8 Normal LSP(s) Found.
 TOTAL: 4 Liberal LSP(s) Found.
 TOTAL: 0 Frr LSP(s) Found.
 A '*' before an LSP means the LSP is not established
 A '*' before a Label means the USCB or DSCB is stale
 A '*' before a UpstreamPeer means the session is stale
 A '*' before a DS means the session is stale
 A '*' before a NextHop means the LSP is FRR LSP
```

下面是在 LSRA 上配置以上 lsp-trigger 策略后，在 LSRB 上执行 **display mpls ldp lsp** 命令后的输出信息。从中可以看出，在 LSRB 上受影响的仅是基于 LSRA 上直连的 10.10.1.1/32 的 LSP 都没有了，其他的 LSP 没受影响，究其原因是 FEC 在 LSRA 上都被禁止建立 LSP 了，自然不会再向其上游进行标签映射消息通告了，从而使得所有下游都不会为其建立 LSP。

```
<AR2>display mpls ldp lsp

 LDP LSP Information
 -------------------------------------------------------------------------------
 DestAddress/Mask   In/OutLabel    UpstreamPeer    NextHop         OutInterface
 -------------------------------------------------------------------------------

 10.10.1.2/32       3/NULL         10.10.1.1       127.0.0.1       InLoop0
 10.10.1.2/32       3/NULL         10.10.1.3       127.0.0.1       InLoop0
*10.10.1.2/32       Liberal/1024                   DS/10.10.1.1
*10.10.1.2/32       Liberal/1025                   DS/10.10.1.3
 10.10.1.3/32       NULL/3         -               10.2.1.2        GE2/0/0
 10.10.1.3/32       1026/3         10.10.1.1       10.2.1.2        GE2/0/0
 10.10.1.3/32       1026/3         10.10.1.3       10.2.1.2        GE2/0/0
*10.10.1.3/32       Liberal/1025                   DS/10.10.1.1
 -------------------------------------------------------------------------------
 TOTAL: 5 Normal LSP(s) Found.
 TOTAL: 3 Liberal LSP(s) Found.
 TOTAL: 0 Frr LSP(s) Found.
 A '*' before an LSP means the LSP is not established
```

```
A '*' before a Label means the USCB or DSCB is stale
A '*' before a UpstreamPeer means the session is stale
A '*' before a DS means the session is stale
A '*' before a NextHop means the LSP is FRR LSP
```

3.3.11 Transit LSP 建立的触发策略配置示例

如图 3-15 所示的 MPLS 网络中，各 LSR 接口上使能 MPLS LDP 后，LDP LSP 将自动建立。由于网络规模比较大，会使各 LSR 建立大量的 LSP，现要求能够控制 LSP 建立的数量，从而减少系统资源的浪费。

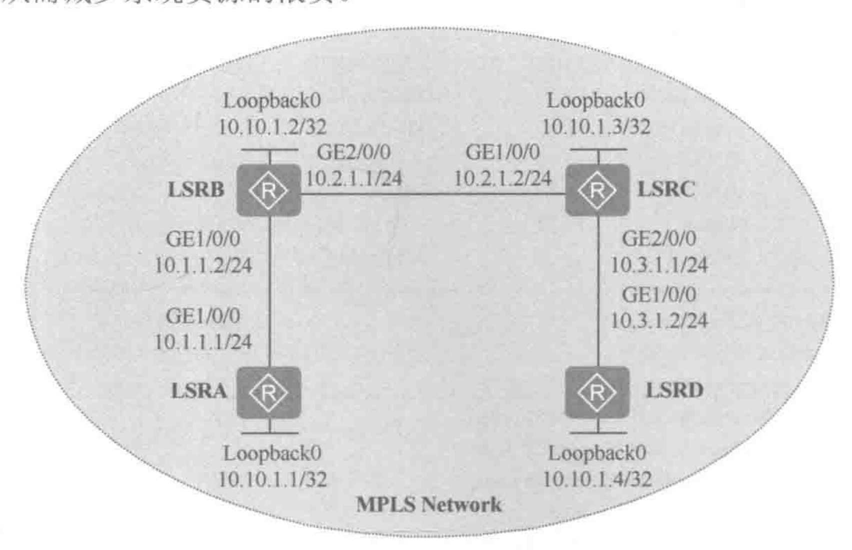


图 3-15 Transit LSP 建立的触发策略配置示例的拓扑结构

1. 基本配置思路分析

本示例的要求与 3.3.10 节介绍的配置示例的要求是一样的，只不过本示例要采用另外一种 LSP 建立过滤方式，即通过在下游配置 propagate mapping 策略，控制下游向上游发送标签映射消息的方式，使得上游边缘设备建立的 LDP LSP 数量减少。本示例是通过在 LSRB 上配置只允许通过过滤条件的路由 10.10.1.4/32 的 FEC 建立 LSP 的 propagate mapping 策略来实现，以减少 LSRA 上的 LSP 的数量，从而节约网络资源。

本示例的基本配置思路与 3.3.10 节介绍的配置示例一样，不同的只是最后一项实现 LSP 建立过滤的手段不同，具体如下。

（1）配置各 LSR 各接口（包括 Loopback 接口）的 IP 地址。

（2）配置各 LSR 的路由，实现网络互通。本示例采用 OSPF 路由协议。

（3）在各 LSR 上使有全局、公网接口的 MPLS 和 LDP 功能。

（4）在 LSRB 上配置 propagate mapping 策略，假设只允许 10.10.1.3/32 的标签映射消息向 LSRA 发送（LSRB 本地网段的标签映射消息不能限制）。

2. 具体配置步骤

（1）配置各接口（包括 Loopback0 接口）的 IP 地址。

\# LSRA 上的配置。

```
<Huawei> system-view
[Huawei] sysname LSRA
[LSRA] interface loopback 0
```

```
[LSRA-LoopBack0] ip address 10.10.1.1 32
[LSRA-LoopBack0] quit
[LSRA] interface gigabitethernet 1/0/0
[LSRA-GigabitEthernet1/0/0] ip address 10.1.1.1 24
[LSRA-GigabitEthernet1/0/0] quit
```

LSRB 上的配置。

```
<Huawei> system-view
[Huawei] sysname LSRB
[LSRB] interface loopback 0
[LSRB-LoopBack0] ip address 10.10.1.2 32
[LSRB-LoopBack0] quit
[LSRB] interface gigabitethernet 1/0/0
[LSRB-GigabitEthernet1/0/0] ip address 10.1.1.2 24
[LSRB-GigabitEthernet1/0/0] quit
[LSRB] interface gigabitethernet 2/0/0
[LSRB-GigabitEthernet2/0/0] ip address 10.2.1.1 24
[LSRB-GigabitEthernet2/0/0] quit
```

LSRC 上的配置。

```
<Huawei> system-view
[Huawei] sysname LSRC
[LSRC] interface loopback 0
[LSRC-LoopBack0] ip address 10.10.1.3 32
[LSRC-LoopBack0] quit
[LSRC] interface gigabitethernet 1/0/0
[LSRC-GigabitEthernet1/0/0] ip address 10.2.1.2 24
[LSRC-GigabitEthernet1/0/0] quit
[LSRC] interface gigabitethernet 2/0/0
[LSRC-GigabitEthernet2/0/0] ip address 10.3.1.1 24
[LSRC-GigabitEthernet2/0/0] quit
```

LSRD 上的配置。

```
<Huawei> system-view
[Huawei] sysname LSRD
[LSRD] interface loopback 0
[LSRD-LoopBack0] ip address 10.10.1.4 32
[LSRD-LoopBack0] quit
[LSRD] interface gigabitethernet 1/0/0
[LSRD-GigabitEthernet1/0/0] ip address 10.3.1.2 24
[LSRD-GigabitEthernet1/0/0] quit
```

（2）配置 OSPF 协议发布各节点公网接口所连网段和 LSR ID 的主机路由，加入 OSPF 路由进程 1、区域 0 中。

LSRA 上的配置。

```
[LSRA] ospf 1
[LSRA-ospf-1] area 0
[LSRA-ospf-1-area-0.0.0.0] network 10.10.1.1 0.0.0.0
[LSRA-ospf-1-area-0.0.0.0] network 10.1.1.0 0.0.0.255
[LSRA-ospf-1-area-0.0.0.0] quit
[LSRA-ospf-1] quit
```

LSRB 上的配置。

```
[LSRB] ospf 1
[LSRB-ospf-1] area 0
[LSRB-ospf-1-area-0.0.0.0] network 10.10.1.2 0.0.0.0
[LSRB-ospf-1-area-0.0.0.0] network 10.1.1.0 0.0.0.255
```

```
[LSRB-ospf-1-area-0.0.0.0] network 10.2.1.0 0.0.0.255
[LSRB-ospf-1-area-0.0.0.0] quit
[LSRB-ospf-1] quit
```

LSRC 上的配置。

```
[LSRC] ospf 1
[LSRC-ospf-1] area 0
[LSRC-ospf-1-area-0.0.0.0] network 10.10.1.3 0.0.0.0
[LSRC-ospf-1-area-0.0.0.0] network 10.2.1.0 0.0.0.255
[LSRC-ospf-1-area-0.0.0.0] network 10.3.1.0 0.0.0.255
[LSRC-ospf-1-area-0.0.0.0] quit
[LSRC-ospf-1] quit
```

LSRD 上的配置。

```
[LSRD] ospf 1
[LSRD-ospf-1] area 0
[LSRD-ospf-1-area-0.0.0.0] network 10.10.1.4 0.0.0.0
[LSRD-ospf-1-area-0.0.0.0] network 10.3.1.0 0.0.0.255
[LSRD-ospf-1-area-0.0.0.0] quit
[LSRD-ospf-1] quit
```

（3）配置 MPLS LDP。以各自的 Loopback0 接口 IP 地址作为它们的 MPLS LSR ID，在全局及接口上使能 MPLS 和 LDP 功能。

LSRA 上的配置。

```
[LSRA] mpls lsr-id 10.10.1.1
[LSRA] mpls
[LSRA-mpls] quit
[LSRA] mpls ldp
[LSRA-mpls-ldp] quit
[LSRA] interface gigabitethernet 1/0/0
[LSRA-GigabitEthernet1/0/0] mpls
[LSRA-GigabitEthernet1/0/0] mpls ldp
[LSRA-GigabitEthernet1/0/0] quit
```

LSRB 上的配置。

```
[LSRB] mpls lsr-id 10.10.1.2
[LSRB] mpls
[LSRB-mpls] quit
[LSRB] mpls ldp
[LSRB-mpls-ldp] quit
[LSRB] interface gigabitethernet 1/0/0
[LSRB-GigabitEthernet1/0/0] mpls
[LSRB-GigabitEthernet1/0/0] mpls ldp
[LSRB-GigabitEthernet1/0/0] quit
[LSRB] interface gigabitethernet 2/0/0
[LSRB-GigabitEthernet2/0/0] mpls
[LSRB-GigabitEthernet2/0/0] mpls ldp
[LSRB-GigabitEthernet2/0/0] quit
```

LSRC 上的配置。

```
[LSRC] mpls lsr-id 10.10.1.3
[LSRC] mpls
[LSRC-mpls] quit
[LSRC] mpls ldp
[LSRC-mpls-ldp] quit
[LSRC] interface gigabitethernet 1/0/0
[LSRC-GigabitEthernet1/0/0] mpls
```

```
[LSRC-GigabitEthernet1/0/0] mpls ldp
[LSRC-GigabitEthernet1/0/0] quit
[LSRC] interface gigabitethernet 2/0/0
[LSRC-GigabitEthernet2/0/0] mpls
[LSRC-GigabitEthernet2/0/0] mpls ldp
[LSRC-GigabitEthernet2/0/0] quit
```

#　LSRD 上的配置。

```
[LSRD] mpls lsr-id 10.10.1.4
[LSRD] mpls
[LSRD-mpls] quit
[LSRD] mpls ldp
[LSRD-mpls-ldp] quit
[LSRD] interface gigabitethernet 1/0/0
[LSRD-GigabitEthernet1/0/0] mpls
[LSRD-GigabitEthernet1/0/0] mpls ldp
[LSRD-GigabitEthernet1/0/0] quit
```

以上配置完成后，在各节点上执行 **display mpls ldp lsp** 命令，可以看到它们 LDP LSP 的建立情况。以下是在 LSRA 上执行该命令的输出示例，按照缺省配置建立了所有 32 位掩码主机路由对应的 LSP。

```
[LSRA] display mpls ldp lsp

 LDP LSP Information
 -------------------------------------------------------------------------------
 DestAddress/Mask   In/OutLabel    UpstreamPeer    NextHop         OutInterface
 -------------------------------------------------------------------------------
 10.10.1.1/32       3/NULL         10.10.1.2       127.0.0.1       InLoop0
*10.10.1.1/32       Liberal/3                      DS/10.10.1.2
 10.10.1.2/32       NULL/3         -               10.1.1.2        GE1/0/0
 10.10.1.2/32       1024/3         10.10.1.2       10.1.1.2        GE1/0/0
 10.10.1.3/32       NULL/1025      -               10.1.1.2        GE1/0/0
 10.10.1.3/32       1022/1025      10.10.1.2       10.1.1.2        GE1/0/0
 10.10.1.4/32       NULL/4118      -               10.1.1.2        GE1/0/0
 10.10.1.4/32       4105/4118      10.10.1.2       10.1.1.2        GE1/0/0
 -------------------------------------------------------------------------------
 TOTAL: 7 Normal LSP(s) Found.
 TOTAL: 1 Liberal LSP(s) Found.
 TOTAL: 0 Frr LSP(s) Found.
 A '*' before an LSP means the LSP is not established
 A '*' before a Label means the USCB or DSCB is stale
 A '*' before a UpstreamPeer means the session is state
 A '*' before a DS means the session is state
 A '*' before a NextHop means the LSP is FRR LSP
```

（4）在 LSRB 上配置 IP 前缀列表，并将此 IP 前缀列表对用于 propagate mapping 策略，对所发送的标签映射消息进行过滤。

在 LSRB 配置 IP 前缀列表，只允许 LSRD 上的 10.10.1.4/32 在 LSRB 建立 Transit LSP，这样一来也就限制了仅允许 LSRB 向上游发送 Transit 类型（同时携带入标签和出标签）的 10.10.1.4/32 的标签映射消息。

```
[LSRB]ip ip-prefix FilterOnTransit permit 10.10.1.4 32
[LSRB] mpls ldp
[LSRB-mpls-ldp] propagate mapping for ip-prefix FilterOnTransit
[LSRB-mpls-ldp] quit
```

以上配置完成后，再在 LSRA 上执行 **display mpls ldp lsp** 命令，可以看到其上面的 LDP LSP 建立情况。

```
[LSRA] display mpls ldp lsp

 LDP LSP Information
 -------------------------------------------------------------------------------
 DestAddress/Mask   In/OutLabel    UpstreamPeer    NextHop         OutInterface
 -------------------------------------------------------------------------------
 10.10.1.1/32       3/NULL         10.10.1.2       127.0.0.1       InLoop0
 10.10.1.2/32       NULL/3         -               10.1.1.2        GE1/0/0
 10.10.1.2/32       1024/3         10.10.1.2       10.1.1.2        GE1/0/0
 10.10.1.4/32       NULL/4118      -               10.1.1.2        GE1/0/0
 10.10.1.4/32       4105/4118      10.10.1.2       10.1.1.2        GE1/0/0
 -------------------------------------------------------------------------------
 TOTAL: 5 Normal LSP(s) Found.
 TOTAL: 0 Liberal LSP(s) Found.
 TOTAL: 0 Frr LSP(s) Found.
 A '*' before an LSP means the LSP is not established
 A '*' before a Label means the USCB or DSCB is stale
 A '*' before a UpstreamPeer means the session is state
 A '*' before a DS means the session is state
 A '*' before a NextHop means the LSP is FRR LSP
```

从以上显示信息可以看到，在 LSRB 配置了 LSP 的控制策略，在 LSRA 上过滤掉了 10.10.1.3/32 的 LDP LSP。因为对于 LSRB 来说，基于 10.10.1.3/32 的标签映射消息是 Transit 类型的，但又不在 propagate mapping 策略许可范围。但在 LSRA 上仍存在一些 FEC 对应的 LSP，其中之一就是 10.10.1.2/32，这是 LSRB 本地的主机路由，向 LSRA 发送的标签映射消息是 Egress 类型的，不是 Transit 类型的，不会被 propagate mapping 策略过滤。

另一个是关于 10.10.1.4/32 的 LDP LSP。因为对于 LSRB 来说，基于 10.10.1.4/32 的标签映射消息是 Transit 类型的，但其在 propagate mapping 策略许可范围内，所以会向 LSRA 发送对应的标签映射消息，在 LSRA 上建立该 FEC 的 LSP。另外，因为 propagate mapping 策略只是过滤 Transit 节点上发送的标签映射消息，所以在 LSRA 上所有不是以 LSRA 为 Ingress 的 LDP LSP 也都仍然会存在，如 LSRA 直连的 10.10.1.1/32 对应的 Egress LSP。

3.3.12 其他 LDP 可选基本功能配置

本节最后来介绍一些其他 LDP 可选基本功能的配置方法，包括 MPLS MTU、MPLS TTL 处理以及禁止向对端分配标签。

1. 配置 MPLS MTU

MTU（最大传输单元）的大小决定了发送端一次能够发送报文的最大字节数，如果 MTU 超过了接收端所能够承受的最大值，或者是超过了发送路径上途经的某台设备所能够承受的最大值，这样就会造成报文分片甚至被丢弃，加重网络传输的负担。所以设备在进行通信之前必须要把 MTU 计算明确，才能保证每次发送的报文都能够畅通无阻地到达接收端，确保报文发送一次成功。

LDP MTU=Min { 所有下游设备通告的 MTU，本机出接口 MTU }。通告方式为，把

计算出来的 LDP MTU 值放在 Label Mapping（标签映射）消息的 MTU TLV 里面，然后把 Label Mapping 消息发送给上游。如果 MTU 发生变动，如本机出接口改变或者配置变更，那么 LSR 就应该再次通过 Label Mapping 消息，把重新计算过的 MTU 通告给其所有上游。而本机出接口 MTU 取值如下。

■ 如果没有配置接口的 MPLS MTU 值，则采用接口的 MTU 值。

■ 如果配置了接口的 MPLS MTU 值，则与接口的 MTU 值比较，采用两者中的较小值作为接口实际生效的 MTU 值。

这样，MPLS 在 Ingress 根据 LDP MTU 来决定 MPLS 转发报文的大小，从而避免在 Ingress 发送的报文较大，导致 Transit 转发失败。

接口 MPLS MTU 的配置方法见表 3-11，一般不用配置。

表 3-11　配置接口 MPLS MTU 的步骤

步骤	命令	说明
1	**system-view** 例如：<Huawei> **system-view**	进入系统视图
2	**mpls ldp** 例如：[Huawei] **mpls ldp**	进入 MPLS LDP 视图
3	**undo mtu-signalling** 例如：[Huawei-mpls-ldp] **undo mtu-signalling**	（二选一）禁止发送标签映射消息时携带 MTU TLV。 缺省情况下，发送标签映射消息时携带华为私有的 MTU TLV。 如果其他厂商设备不支持 MTU TLV，为了实现互通则需要禁止发送标签映射消息时携带 MTU TLV。若已禁止 LSR 发送 MTU TLV，则配置的 MPLS MTU 值不生效
	mtu-signalling apply-tlv 例如：[Huawei-mpls-ldp] **mtu-signalling apply-tlv**	（二选一）配置发送标签映射消息时携带 RFC3988 定义的 MTU TLV。使能或去使能 MTU TLV 发送功能的操作将导致原始 LDP 会话重建，造成 MPLS 业务中断。 缺省情况下，发送标签映射消息时携带华为私有的 MTU TLV。 如果其他厂商设备支持 MTU TLV，为了实现互通则需要使 LSR 发送 RFC3988 中定义的标准 MTU TLV，否则可能导致用户配置的 MPLS MTU 值不生效
4	**quit** 例如：[Huawei-mpls-ldp] **quit**	退回系统视图
5	**interface** *interface-type interface-number* 例如：[Huawei] **interface** gigabitethernet 1/0/0	进入使能了 MPLS 的接口视图
6	**mpls mtu** *mtu* 例如:[Huawei-GigabitEthernet1/0/0] **mpls mtu** 1500	配置接口的 MPLS MTU，取值范围与接口类型相关。 缺省情况下，接口 MPLS 报文的 MTU 等于接口本身的 MTU，可用 **undo mpls mtu** 命令恢复缺省值

2. 配置 MPLS 对 TTL 的处理

MPLS 对 TTL 的处理包括两个方面（这两个方面的详细说明参见第 1 章 1.2.4 节）。

（1）MPLS 对 TTL 的处理模式

在 MPLS VPN 应用中，出于网络安全的考虑，需要隐藏 MPLS 骨干网络的结构。在

这种情况下，对于私网报文，Ingress 上使用 MPLS Pipe 模式。若想反映报文实际经过的路径，则在 Ingress 上使用 MPLS Uniform 模式。

（2）ICMP 响应报文使用的路径

缺省情况下，收到的 MPLS 报文只带一层标签时，LSR 使用 IP 路由返回 ICMP 响应报文；收到的 MPLS 报文包含多层标签时，LSR 使用 LSP 返回 ICMP 响应报文。但是，在 MPLS VPN 中，ASBR（自治系统边界路由器）和 HoVPN（分层 VPN）组网应用中的 SPE（Superstratum PE or Sevice Provider-end PE，上层 PE 或运营商侧 PE），接收到的承载 VPN 报文的 MPLS 报文可能只有一层标签，此时，这些设备上并不存在到达报文发送者的路由，则 LSR 使用 LSP 返回 ICMP 响应报文。

MPLS 对 TTL 的处理方法只需在 Ingress 上或同时包括 Egress 上配置，具体见表 3-12。

表 3-12 配置 MPLS TTL 处理方法的步骤

步骤	命令	说明
1	**system-view** 例如：<Huawei> **system-view**	进入系统视图
配置 MPLS 对 TTL 的处理模式（仅需在 Ingress 节点上配置）		
2	**ttl propagate** 例如：[Huawei] **ttl propagate**	配置 MPLS TTL 的处理模式为 Uniform 模式。 缺省情况下，MPLS 报文中 TTL 传播模式是 Uniform，可用 **undo ttl propagate** 命令配置 MPLS TTL 的处理模式为 Pipe 模式。 **【注意】**配置本命令只影响此后新建立的 LSP，如果需要对之前建立的 LSP 也生效，应执行 **reset mpls ldp** 命令重建 LSP
配置 ICMP 响应报文使用的路径		
3	**mpls** 例如：[Huawei] **mpls**	进入 MPLS 视图
4	**ttl expiration pop** 例如：[Huawei-mpls] **undo ttl expiration pop**	使用 IP 路由返回 ICMP 响应报文。 缺省情况下，对于一层标签的 MPLS TTL 超时报文，将根据本地 IP 路由返回 ICMP 报文，可用 **undo ttl expiration pop** 命令使用 LSP 返回 ICMP 响应报文

3. 禁止向远端对等体分配标签

在以 LDP 作为信令协议的 MPLS L2VPN 应用场景中（包括 Martini 方式的 VLL、PWE3 等），VPN 两端的 PE 之间通常需要建立 LDP 远端会话。这里的远端会话仅用于传递私网标签的 Label-Mapping 消息，因此不需要 LDP 为其分配 LDP 标签。但是，缺省情况下，LDP 会为远端对等体分配普通的 LDP 标签。这将产生很多无用的空闲标签，浪费 LDP 的标签资源。

为了解决上述问题，可以配置禁止向远端对等体分配标签，以节省系统资源。禁止向远端对等体分配标签的配置方式有以下两种。

- 在 LDP 视图下禁止向所有远端对等体分配标签。
- 在指定远端对等体视图下禁止向该对等体分配标签。

以上两种方式的具体配置方法见表 3-13。

表 3-13　配置禁止向远端对等体分配标签的步骤

步骤	命令	说明
1	**system-view** 例如：<Huawei> **system-view**	进入系统视图
方法一：禁止向指定的远端邻居分发公网标签		
2	**mpls ldp remote-peer** *remote-peer-name* 例如：[Huawei] **mpls ldp remote-peer** Hunan	进入 MPLS LDP 远端对等体视图。参数 *remote-peer-name* 用来指定远端对等体名称，字符串形式，不支持空格，不区分大小写，长度范围是 1～32。当输入的字符串两端使用双引号时，可在字符串中输入空格
3	**remote-ip** *ip-address* **pwe3** 例如：[Huawei-mpls-ldp-remote-Hunan] **remote-ip** 10.1.1.1 **pwe3**	配置禁止向指定的远端对等体分发公网标签。参数 *ip-address* 用来指定远端对等体 IP 地址，必须是远端对等体的 LSR ID。LDP LSR ID 和 MPLS LSR ID 不一致时，要使用 LDP LSR ID。 【注意】通过本命令配置远端对等体的 IP 地址后，该 IP 地址不能再作为本地接口的 IP 地址，否则将导致远端会话被中断。 缺省情况下，没有配置 LDP 远端对等体的 IP 地址，可用 **undo remote-ip pwe3** 删除原来的配置
方法二：禁止向所有的远端邻居分发公网标签		
2	**mpls ldp** 例如：[Huawei] **mpls ldp**	进入 MPLS LDP 视图
3	**remote-peer pwe3** 例如：[Huawei-mpls-ldp] **remote-peer pwe3**	配置禁止向所有远端对等体（包括已经存在的远端对等体）分发公网标签。 缺省情况下，允许向所有远端邻居分发公网标签，可用 **undo remote-peer pwe3** 命令恢复缺省配置

3.3.13　禁止向远端对等体分配标签配置示例

如图 3-16 所示，PE1、PE2 和 PE3 由 MPLS 骨干网 P 设备连接，各设备间运行 IS-IS 路由协议。使用公网 LSP 隧道，PE1 分别与 PE2、PE3 建立 LDP 远端会话来传递私网标签信息，在 PE1 和 PE2 之间、PE1 和 PE3 之间建立动态 PW。要求能够控制 LDP 向远端对等体分配标签，以节约系统资源。

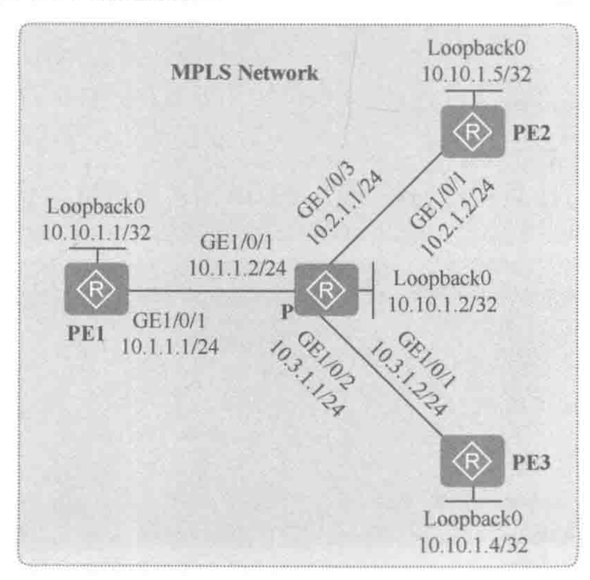


图 3-16　禁止向远端对等体分配标签配置示例的拓扑结构

1. 基本配置思路分析

本示例是希望控制 LDP 向远端对等体分配标签，这时可在 PE 之间配置禁止向远端对等体分配标签策略，禁止 PE1 与 PE2、PE3 间分配普通的 LDP 标签，以节约系统资源。但这项功能是可选的 LDP 基本功能，在配置此功能前还需要先完成 LDP 必选基本功能配置。

本示例的基本配置思路如下：

（1）配置各设备接口（包括 Loopback 接口）IP 地址；

（2）配置各设备间的路由。本示例采用 OSPF 路由协议；

（3）在各设备上全局及公网接口使能 MPLS、LDP 功能；

（4）配置 PE_1 分别与 PE_2、PE_3 之间的远端对等体关系；

（5）配置 PE_1 分别与 PE_2、PE_3 远端对等体之间禁止相互分配标签。

2. 具体配置步骤

（1）配置各设备接口（包括 Loopback 接口）的 IP 地址。

PE1 上的配置。

```
<Huawei> system-view
[Huawei] sysname PE1
[PE1] interface loopback0
[PE1-LoopBack0] ip address 10.10.1.1 32
[PE1-LoopBack0] quit
[PE1] interface gigabitethernet 1/0/1
[PE1-GigabitEthernet1/0/1] ip address 10.1.1.1 24
[PE1-GigabitEthernet1/0/1] quit
```

P 上的配置。

```
<Huawei> system-view
[Huawei] sysname P
[P] interface loopback0
[P-LoopBack0] ip address 10.10.1.2 32
[P-LoopBack0] quit
[P] interface gigabitethernet 1/0/1
[P-GigabitEthernet1/0/1] ip address 10.1.1.2 24
[P-GigabitEthernet1/0/1] quit
[P] interface gigabitethernet 1/0/2
[P-GigabitEthernet1/0/2] ip address 10.3.1.1 24
[P-GigabitEthernet1/0/2] quit
[P] interface gigabitethernet 1/0/3
[P-GigabitEthernet1/0/3] ip address 10.2.1.1 24
[P-GigabitEthernet1/0/3] quit
```

PE2 上的配置。

```
<Huawei> system-view
[Huawei] sysname PE2
[PE2] interface loopback0
[PE2-LoopBack0] ip address 10.10.1.5 32
[PE2-LoopBack0] quit
[PE2] interface gigabitethernet 1/0/1
[PE2-GigabitEthernet1/0/1] ip address 10.2.1.2 24
[PE2-GigabitEthernet1/0/1] quit
```

PE3 上的配置。

```
<Huawei> system-view
```

```
[Huawei] sysname PE3
[PE3] interface loopback0
[PE3-LoopBack0] ip address 10.10.1.4 32
[PE3-LoopBack0] quit
[PE3] interface gigabitethernet 1/0/1
[PE3-GigabitEthernet1/0/1] ip address 10.3.1.2 24
[PE3-GigabitEthernet1/0/1] quit
```

（2）配置 OSPF 协议发布各节点公网接口所连网段和 LSR ID 的主机路由，加入 OSPF 路由进程 1、区域 0 中。

#　PE_1 上的配置。

```
[PE_1] ospf 1
[PE_1-ospf-1] area 0
[PE_1-ospf-1-area-0.0.0.0] network 10.10.1.1 0.0.0.0
[PE_1-ospf-1-area-0.0.0.0] network 10.1.1.0 0.0.0.255
[PE_1-ospf-1-area-0.0.0.0] quit
[PE_1-ospf-1] quit
```

#　P 上的配置。

```
[P] ospf 1
[P-ospf-1] area 0
[P-ospf-1-area-0.0.0.0] network 10.10.1.2 0.0.0.0
[P-ospf-1-area-0.0.0.0] network 10.1.1.0 0.0.0.255
[P-ospf-1-area-0.0.0.0] network 10.2.1.0 0.0.0.255
[P-ospf-1-area-0.0.0.0] network 10.3.1.0 0.0.0.255
[P-ospf-1-area-0.0.0.0] quit
[P-ospf-1] quit
```

#　PE_2 上的配置。

```
[PE_2] ospf 1
[PE_2-ospf-1] area 0
[PE_2-ospf-1-area-0.0.0.0] network 10.10.1.5 0.0.0.0
[PE_2-ospf-1-area-0.0.0.0] network 10.2.1.0 0.0.0.255
[PE_2-ospf-1-area-0.0.0.0] quit
[PE_2-ospf-1] quit
```

#　PE_3 上的配置。

```
[PE_3] ospf 1
[PE_3-ospf-1] area 0
[PE_3-ospf-1-area-0.0.0.0] network 10.10.1.4 0.0.0.0
[PE_3-ospf-1-area-0.0.0.0] network 10.3.1.0 0.0.0.255
[PE_3-ospf-1-area-0.0.0.0] quit
[PE_3-ospf-1] quit
```

（3）使能各节点全局和各设备间相连的公网接口的 MPLS 和 MPLS LDP 功能。

#　PE1 上的配置。

```
[PE1] mpls lsr-id 10.10.1.1
[PE1] mpls
[PE1-mpls] quit
[PE1] mpls ldp
[PE1-mpls-ldp] quit
[PE1] interface gigabitethernet 1/0/1
[PE1-GigabitEthernet1/0/1] mpls
[PE1-GigabitEthernet1/0/1] mpls ldp
[PE1-GigabitEthernet1/0/1] quit
```

#　P 上的配置。

```
[P] mpls lsr-id 10.10.1.2
[P] mpls
[P-mpls] quit
[P] mpls ldp
[P-mpls-ldp] quit
[P] interface gigabitethernet 1/0/1
[P-GigabitEthernet1/0/1] mpls
[P-GigabitEthernet1/0/1] mpls ldp
[P-GigabitEthernet1/0/1] quit
[P] interface gigabitethernet 1/0/2
[P-GigabitEthernet1/0/2] mpls
[P-GigabitEthernet1/0/2] mpls ldp
[P-GigabitEthernet1/0/2] quit
[P] interface gigabitethernet 1/0/3
[P-GigabitEthernet1/0/3] mpls
[P-GigabitEthernet1/0/3] mpls ldp
[P-GigabitEthernet1/0/3] quit
```

PE2 上的配置。

```
[PE2] mpls lsr-id 10.10.1.5
[PE2] mpls
[PE2-mpls] quit
[PE2] mpls ldp
[PE2-mpls-ldp] quit
[PE2] interface gigabitethernet 1/0/1
[PE2-GigabitEthernet1/0/1] mpls
[PE2-GigabitEthernet1/0/1] mpls ldp
[PE2-GigabitEthernet1/0/1] quit
```

PE3 上的配置。

```
[PE3] mpls lsr-id 10.10.1.4
[PE3] mpls
[PE3-mpls] quit
[PE3] mpls ldp
[PE3-mpls-ldp] quit
[PE3] interface gigabitethernet 1/0/1
[PE3-GigabitEthernet1/0/1] mpls
[PE3-GigabitEthernet1/0/1] mpls ldp
[PE3-GigabitEthernet1/0/1] quit
```

上述配置完成后，相邻节点之间应该建立起 LDP 会话以及公网 LSP。在各节点上执行 **display mpls ldp session** 命令可以看到设备间的 LDP 会话状态为“Operational”，表示 LDP 会话建立成功。以下是在 PE1 上执行该命令的输出示例，显示其仅存在一个与 P 之间的 LDP 会话，且建立状态为“Operational”，表示会话建立成功。

```
[PE1] display mpls ldp session

 LDP Session(s) in Public Network
 Codes: LAM(Label Advertisement Mode), SsnAge Unit(DDDD:HH:MM)
 A '*' before a session means the session is being deleted.
 ------------------------------------------------------------------------------
 PeerID             Status        LAM   SsnRole   SsnAge        KASent/Rcv
 ------------------------------------------------------------------------------
 10.10.1.2:0        Operational   DU    Passive   0000:00:01    6/6
 ------------------------------------------------------------------------------
 TOTAL: 1 session(s) Found.
```

执行 **display mpls ldp lsp** 命令可以看到建立的 LSP 情况和标签的分配情况。以下是在 PE_1 上执行该命令的输出示例，从中可以看出，MPLS 域中所有 32 位掩码主机路由都建立了 LDP LSP。

```
[PE1] display mpls ldp lsp

 LDP LSP Information
 -------------------------------------------------------------------------------
 DestAddress/Mask    In/OutLabel      UpstreamPeer     NextHop          OutInterface
 -------------------------------------------------------------------------------
 10.10.1.1/32        3/NULL           10.10.1.2        127.0.0.1        InLoop0
*10.10.1.1/32        Liberal/1025                      DS/10.10.1.2
 10.10.1.2/32        NULL/3           -                10.1.1.2         GE1/0/1
 10.10.1.2/32        1024/3           10.10.1.2        10.1.1.2         GE1/0/1
 10.10.1.4/32        NULL/1024        -                10.1.1.2         GE1/0/1
 10.10.1.4/32        1025/1024        10.10.1.2        10.1.1.2         GE1/0/1
 10.10.1.5/32        NULL/1026        -                10.1.1.2         GE1/0/1
 10.10.1.5/32        1022/1026        10.10.1.2        10.1.1.2         GE1/0/1
 -------------------------------------------------------------------------------
 TOTAL: 7 Normal LSP(s) Found.
 TOTAL: 1 Liberal LSP(s) Found.
 TOTAL: 0 Frr LSP(s) Found.
 A '*' before an LSP means the LSP is not established
 A '*' before a Label means the USCB or DSCB is stale
 A '*' before a UpstreamPeer means the session is stale
 A '*' before a DS means the session is stale
 A '*' before a NextHop means the LSP is FRR LSP
```

（4）在 PE_1 分别与 PE_2、PE_3 之间建立 MPLS LDP 远端对等体关系。

PE1 上的配置。

```
[PE1] mpls ldp remote-peer PE2
[PE1-mpls-ldp-remote-pe2] remote-ip 10.10.1.5
[PE1-mpls-ldp-remote-pe2] quit
[PE1] mpls ldp remote-peer PE3
[PE1-mpls-ldp-remote-pe3] remote-ip 10.10.1.4
[PE1-mpls-ldp-remote-pe3] quit
```

PE2 上的配置。

```
[PE2] mpls ldp remote-peer PE1
[PE2-mpls-ldp-remote-pe1] remote-ip 10.10.1.1
[PE2-mpls-ldp-remote-pe1] quit
```

PE3 上的配置。

```
[PE3] mpls ldp remote-peer PE1
[PE3-mpls-ldp-remote-pe1] remote-ip 10.10.1.1
[PE3-mpls-ldp-remote-pe1] quit
```

上述配置完成后，各 PE 节点之间应该建立起远端 LDP 会话。在各节点上执行 **display mpls ldp session** 命令可以看到各 PE 设备间建立的远端会话。

以下是在 PE_1 上执行该命令的输出示例，从中可以看出，除了原来与 P 之间建立的本地 LDP 会话外又多了两条分别与 PE_2 和 PE_3 之间的远端会话，且状态均为“Operational”，表示会话建立成功。这样就证明前面的 LDP 远端会话的配置是正确的。

```
[PE1]display mpls ldp session

 LDP Session(s) in Public Network
 Codes: LAM(Label Advertisement Mode), SsnAge Unit(DDDD:HH:MM)
 A '*' before a session means the session is being deleted.
 ------------------------------------------------------------------------------
 PeerID                Status          LAM   SsnRole   SsnAge        KASent/Rcv
 ------------------------------------------------------------------------------
 10.10.1.2:0           Operational DU   Passive   0000:00:18   75/75
 10.10.1.4:0           Operational DU   Passive   0000:00:10   43/43
 10.10.1.5:0           Operational DU   Passive   0000:00:12   50/50
 ------------------------------------------------------------------------------
 TOTAL: 3 session(s) Found.
```

再次执行 **display mpls ldp lsp** 命令可以看到，各 PE 设备都给自己的远端对等体分配了 Liberal 标签，但实际上这些标签在 MPLS L2VPN 应用中是空闲无用的，且占用了大量系统资源。

以下是在 PE_1 上执行该命令的输出示例，相比在没有配置 PE 设备间 LDP 远端会话前所建立的 LSP，多出了一些分别向远端对等体 PE_1 和 PE_2 为各 FEC 分配的 Liberal 标签（参见输出信息中的粗体字部分），这些都是没有建立成功的 LSP。

```
[PE1] display mpls ldp lsp

 LDP LSP Information
 ------------------------------------------------------------------------------
 DestAddress/Mask   In/OutLabel    UpstreamPeer    NextHop         OutInterface
 ------------------------------------------------------------------------------
 10.10.1.1/32       3/NULL         10.10.1.2       127.0.0.1       InLoop0
 10.10.1.1/32       3/NULL         10.10.1.5       127.0.0.1       InLoop0
 10.10.1.1/32       3/NULL         10.10.1.4       127.0.0.1       InLoop0
*10.10.1.1/32       Liberal/1025                   DS/10.10.1.2
*10.10.1.1/32       Liberal/1024                   DS/10.10.1.5
*10.10.1.1/32       Liberal/1025                   DS/10.10.1.4
 10.10.1.2/32       NULL/3         -               10.1.1.2        GE1/0/1
 10.10.1.2/32       1024/3         10.10.1.2       10.1.1.2        GE1/0/1
 10.10.1.2/32       1024/3         10.10.1.5       10.1.1.2        GE1/0/1
 10.10.1.2/32       1024/3         10.10.1.4       10.1.1.2        GE1/0/1
 10.10.1.4/32       NULL/1024      -               10.1.1.2        GE1/0/1
 10.10.1.4/32       1025/1024      10.10.1.2       10.1.1.2        GE1/0/1
 10.10.1.4/32       1025/1024      10.10.1.5       10.1.1.2        GE1/0/1
 10.10.1.4/32       1025/1024      10.10.1.4       10.1.1.2        GE1/0/1
*10.10.1.4/32       Liberal/1026                   DS/10.10.1.5
*10.10.1.4/32       Liberal/3                      DS/10.10.1.4
 10.10.1.5/32       NULL/1026      -               10.1.1.2        GE1/0/1
 10.10.1.5/32       1022/1026      10.10.1.2       10.1.1.2        GE1/0/1
 10.10.1.5/32       1022/1026      10.10.1.5       10.1.1.2        GE1/0/1
 10.10.1.5/32       1022/1026      10.10.1.4       10.1.1.2        GE1/0/1
*10.10.1.5/32       Liberal/3                      DS/10.10.1.5
*10.10.1.5/32       Liberal/1026                   DS/10.10.1.4
 ------------------------------------------------------------------------------
 TOTAL: 15 Normal LSP(s) Found.
 TOTAL: 9 Liberal LSP(s) Found.
 TOTAL: 0 Frr LSP(s) Found.
 A '*' before an LSP means the LSP is not established
 A '*' before a Label means the USCB or DSCB is stale
```

```
A '*' before a UpstreamPeer means the session is stale
A '*' before a DS means the session is stale
A '*' before a NextHop means the LSP is FRR LSP
```

如果在 PE_2 和 PE_3 上执行 **display mpls ldp lsp** 命令同样可以看到，它们为 PE_1 基于各 FEC 分配的 Liberal 标签。

（5）在 PE_1 分别与 PE_2、PE_3 之间配置禁止向远端对等体分配标签。

PE1 上的配置。

```
[PE1] mpls ldp remote-peer PE2
[PE1-mpls-ldp-remote-pe2] remote-ip 10.10.1.5 pwe3
[PE1-mpls-ldp-remote-pe2] quit
[PE1] mpls ldp remote-peer PE3
[PE1-mpls-ldp-remote-pe3] remote-ip 10.10.1.4 pwe3
[PE1-mpls-ldp-remote-pe3] quit
```

PE2 上的配置。

```
[PE2] mpls ldp remote-peer PE1
[PE2-mpls-ldp-remote-pe1] remote-ip 10.10.1.1 pwe3
[PE2-mpls-ldp-remote-pe1] quit
```

PE3 上的配置。

```
[PE3] mpls ldp remote-peer PE1
[PE3-mpls-ldp-remote-pe1] remote-ip 10.10.1.1 pwe3
[PE3-mpls-ldp-remote-pe1] quit
```

上述配置完成后，相邻节点之间应该 LDP 远端会话所分配的 Liberal 标签将会被禁止。在各 PE 节点上执行 **display mpls ldp lsp** 命令可以看到配置禁止向远端对等体分配标签后的 LSP 建立情况。以下是在 PE1 上执行该命令的输出示例。

```
[PE1] display mpls ldp lsp

 LDP LSP Information
 -------------------------------------------------------------------------------
 DestAddress/Mask   In/OutLabel    UpstreamPeer    NextHop        OutInterface
 -------------------------------------------------------------------------------
 10.10.1.1/32       3/NULL          10.10.1.2      127.0.0.1       InLoop0
*10.10.1.1/32       Liberal/1025                   DS/10.10.1.2
 10.10.1.2/32       NULL/3          -                10.1.1.2        GE1/0/1
 10.10.1.2/32       1024/3         10.10.1.2       10.1.1.2       GE1/0/1
 10.10.1.4/32       NULL/1024       -                10.1.1.2        GE1/0/1
 10.10.1.4/32       1025/1024      10.10.1.2       10.1.1.2       GE1/0/1
 10.10.1.5/32       NULL/1026       -                10.1.1.2        GE1/0/1
 10.10.1.5/32       1022/1026      10.10.1.2       10.1.1.2       GE1/0/1
 -------------------------------------------------------------------------------
 TOTAL: 7 Normal LSP(s) Found.
 TOTAL: 1 Liberal LSP(s) Found.
 TOTAL: 0 Frr LSP(s) Found.
 A '*' before an LSP means the LSP is not established
 A '*' before a Label means the USCB or DSCB is stale
 A '*' before a UpstreamPeer means the session is stale
 A '*' before a DS means the session is stale
 A '*' before a NextHop means the LSP is FRR LSP
```

如果在 PE_2 和 PE_3 上执行 **display mpls ldp lsp** 命令同样可以看到，它们原来为 PE_1 基于各 FEC 分配的 Liberal 标签也全部没有了。LSP 的建立情况又恢复到了只有本地会话的情况。

3.4 LDP LSP 建立典型故障排除

在动态 LDP LSP 建立过程中可能因为配置错误而出现一些故障，典型故障包括 LDP 会话振荡、LDP 会话 Down、LDP LSP Down、无法建立跨域 LSP，下面具体介绍排除方法。

1. LDP 会话振荡故障排除

LDP 会话振荡是指节点间的 LDP 会话建立时而成功，时而失败，这主要是对 LDP GR 定时器、LDP MTU、LDP 认证、LDP Keepalive 定时器、LDP 传输地址的配置进行新增、修改或删除造成的。具体的排除步骤如下。

（1）在各节点的 LDP 视图下执行 **display this** 命令，查看是否进行了 LDP GR 或 LDP MTU 配置。如果显示信息中包含了以下配置，则表示进行了 LDP GR 配置。

```
mpls ldp
 graceful-restart
```

如果显示信息中包含了以下配置，表示进行了 LDP MTU 配置。

```
mpls ldp
 mtu-signalling apply-tlv
```

如果显示信息中包含（具体数值依据实际情况而异）配置，则表示进行了 LDP 认证配置。

```
mpls ldp
 md5-password cipher 2.2.2.2 @%@%7I$3/^8`u"M|%hKXui~5kO4U@%@%
```

或

```
mpls ldp
 authentication key-chain peer 2.2.2.2 name kc1
```

（2）在节点公网接口视图下执行 **display this** 命令，查看是否执行了 LDP Keepalive 定时器或 LDP 传输地址的配置。

如果显示信息中包含（具体数值依据实际情况而异）以下配置，则表示进行了 LDP Keepalive 定时器配置。

```
mpls ldp
mpls ldp timer keepalive-hold 30
```

如果显示信息中包含（具体数值依据实际情况而异）以下配置，则表示进行了 LDP 传输地址配置。

```
mpls ldp
mpls ldp transport-address interface
```

（3）如果进行了上述配置，请等待 10s，等待 LDP 会话稳定。

2. LDP 会话 Down 故障排除

如果在配置 LDP 会话后发现 LDP 会话状态为 Down，则可按以下步骤进行排除。

（1）在对应节点的公网接口视图下执行 **display this** 命令，查看接口是否被关闭了。如果接口被 Shutdown，请在接口下执行 **undo shutdown** 命令启动接口。

（2）检查是否执行了取消 MPLS 相关配置的命令。

在对应节点上执行 **display current-configuration** 命令，查看是否执行了取消 MPLS

相关配置的命令。

如果显示信息中没有包含以下配置，表示取消了 MPLS 的配置。

```
mpls
```

如果显示信息中没有包含以下配置，则表示取消了 MPLS LDP 的配置。

```
mpls ldp
```

如果显示信息中没有包含以下配置，则表示删除了 LDP 远端会话的配置。

```
mpls ldp remote-peer
```

如果执行了取消 MPLS 相关配置的命令，请执行相应的配置命令恢复被取消的配置。

3. LDP LSP Down 故障排除

如果在 LDP LSP 建立配置完成后发现 LDP LSP 的状态为 Down，则可按以下步骤进行故障排除。

（1）在各节点上执行 **display mpls ldp session** 命令，查看显示信息的 **Status** 字段，检查 LDP 会话是否正常建立。如果该字段显示为 **Operational**，则表示 LDP 会话已建立并处于 Up 状态。如果该字段显示不是 **Operational**，则表示 LDP 会话没有被正常建立。

如果 LDP 会话没有正常被建立，请参见前面介绍的“LDP 会话 Down”故障排除方法继续定位。

（2）在各节点的 MPLS 视图下执行 **display this** 命令，检查是否配置了 LSP 建立策略。如果显示信息中有以下配置（具体数值依据实际情况而异），则需要检查 IP 前缀策略 abc 中是否屏蔽了相关 LSP。

```
lsp-trigger ip-prefix abc
```

（3）在各节点的 MPLS LDP 视图下执行 **display this** 命令，如果显示信息中有以下配置（具体数值依据实际情况而异），则需要检查 IP 前缀策略 abc 中是否屏蔽了相关 LSP。

```
propagate mapping for ip-prefix abc
```

（4）在各节点的系统视图下执行 **display ip ip-prefix** 命令，如果显示信息中有以下配置（具体数值依据实际情况而异），则表示只允许为 10.1.1.1/32、10.2.2.2/32 两个路由建立 LSP。

```
index: 10          permit   10.1.1.1/32
index: 20          permit   10.2.2.2/32
```

（5）如果配置了以上策略，请在策略中增加 LSP 对应的路由信息。

4. 无法建立跨域 LSP 故障排除

如果在配置 LDP 跨域扩展后无法建立跨域 LSP，则可按以下步骤进行故障排除。

（1）在各节点上执行 **display mpls ldp** 命令，查看显示信息的 **Longest-match** 字段，检查是否已配置了 LDP 跨域扩展功能。如果该字段显示为 **On**，则表示使能了 LDP 跨域扩展功能。如果该字段显示为 **Off**，则表示没有使能 LDP 跨域扩展功能。

如果没有使能 LDP 跨域扩展功能，请执行 **longest-match** 命令使能 LDP 跨域扩展功能。

（2）在各节点上执行 **display mpls ldp session** 命令，查看显示信息的 **Status** 字段，检查 LDP 会话是否正常建立。如果该字段的显示为 **Operational**，则表示 LDP 会话已建立并处于 Up 状态。如果该字段显示为非 **Operational** 或者没有会话信息显示，则表示 LDP 会话没有被正常建立。

如果 LDP 会话没有被正常建立，请参见前面介绍的“LDP 会话 Down”故障排除方

法，继续定位。

（3）在各节点上检查 LDP 会话是否与路由匹配。

■ 执行 **display ip routing-table** 命令，记录 **NextHop** 和 **Interface** 字段。

■ 执行 **display mpls ldp session verbose** 命令，记录 **Addresses received from peer** 字段。

■ 执行 **display mpls ldp peer** 命令，记录 **DiscoverySource** 字段。

如果 NextHop 字段的信息包含在 **Addresses received from peer** 字段中，并且 **Interface** 字段信息和 **DiscoverySource** 字段信息相同，则表示 LDP 会话与路由匹配。

如果 LDP 会话和路由不匹配，请参见前面介绍的“LDP LSP Down”故障排除方法继续定位。



第4章 MPLS LDP扩展功能配置与管理



第 3 章介绍了 LDP LSP 中的一些基本功能配置与管理方法，第 4 章再来专门介绍 LDP LSP 中的一些扩展功能的配置与管理方法。这些扩展功能主要包括 LDP 跨域连接，基于 BFD 的 LDP LSP 检测，LDP 与静态路由、IGP 路由联动，以及 LDP FRR、LDP GR 和 LDP 邻居建立的安全机制等。

LDP 跨域连接是指跨越多个 IGP 区域的 LDP LSP，基于 BFD 的 LDP LSP 检测，LDP 与静态路由、IGP 路由联动，以及 LDP FRR、LDP GR 都是为了提高通过 LDP LSP 进行数据传输的可靠性的一些措施，而 LDP 安全机制是为了提高 LDP 邻居建立的安全性。

4.1 配置 LDP 跨域扩展

当 MPLS 骨干网规模比较大时，通常需要部署多个 IGP 区域（如采用 OSPF、IS-IS 路由时）来达到灵活部署和快速收敛的目的。在这种情况下，IGP 区域间进行路由通告时，为了避免路由数量多而引起对资源的过多占用，区域边界路由器（ABR）需要将区域内路由进行聚合，向邻居设备通过区域内的聚合路由。

这时我们会想到，一个区域中的设备如何为其他区域中设备所连接的网段建立 LSP 呢？因为本区域接收不到其他区域各网段的明细路由，只有一个区域聚合路由。而且缺省情况下，LDP 在建立 LSP 的时候，只会在路由表中查找与收到的标签映射消息中携带的 FEC 精确匹配的路由，对于聚合路由，LDP 只能建立 Liberal LSP（自由保持类的 LSP，是一种没有建立成功的 LSP），无法建立跨越 IGP 区域的 LDP LSP。

这时就要依靠 LDP 跨域扩展功能了。LDP 跨域扩展通过使能 LDP 按最长匹配原则（这样一来不一定非要与具体的明细路由匹配）查找路由，使 LDP 能够依据聚合后的路由建立起跨越多个 IGP 区域的 LDP LSP。

如图 4-1 所示，MPLS/IP 骨干网中存在 Area10 和 Area20 两个 IS-IS 区域。在 Area10 区域边缘的 LSR_2 的路由表中，存在到 LSR_3 和 LSR_4 上针对两个 Loopback 接口的两条主机路由，为了避免路由数量多而引起对资源的过多占用，在 LSR_2 上通过 IS-IS 路由协议将这两条路由聚合为 1.3.0.0/24 发送到 Area20 区域。

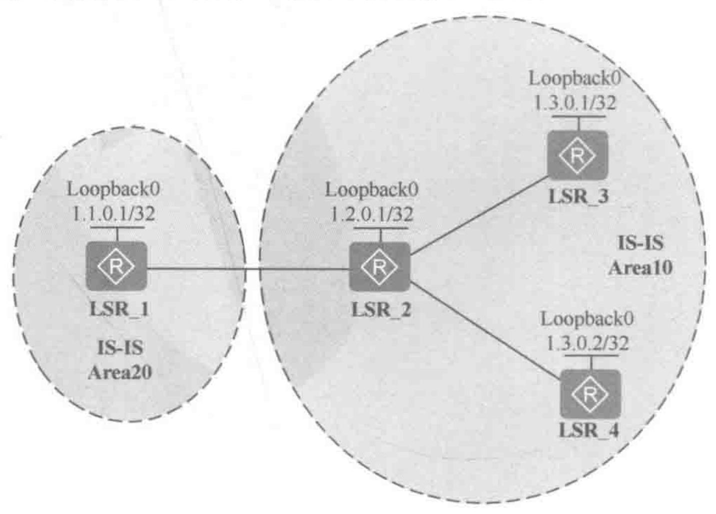


图 4-1 LDP 跨域扩展示例

LDP 在建立 LSP 的时候，会在路由表中查找与收到的标签映射消息中携带的 FEC 精确匹配的路由。对于图 4-1 中的情况，LSR_1 的路由表中针对 Area10 中网段，只有 1.3.0.0/24 这条聚合后的路由，而没有 32 位的主机路由，却对应了两个 FEC：1.3.0.1/32、1.3.0.2/32。

对于聚合路由，LDP 只能建立 Liberal LSP（自由保持 LSP，属于没建立成功的 LSP），无法建立跨越 IGP 区域的 LDP LSP，以至于无法提供必要的骨干网隧道。因此，在 LSR_1 上需要按照最长匹配原则查找能标签映射消息中所携带的两个 FEC：1.3.0.1/32、1.3.0.2/32 匹配的路由来建立 LSP。现在 LSR_1 的路由表中，已经存在聚合路由 1.3.0.0/24。当 LSR_1 收到 Area10 区域的标签映射消息时（例如携带的 FEC 为 1.3.0.1/32），按照最长匹配原则的查找方式，LSR_1 能够找到聚合路由 1.3.0.0/24 的信息，即把该路由的出接口和下一跳作为到达 FEC 1.3.0.1/32 的出接口和下一跳。这样，LDP 就为 1.3.0.1/32 网段建立跨越 IGP 区域的 LDP LSP 了。1.3.0.12/32 网段的 LDP LSP 建立过程一样，其出接口和下一跳与 1.3.0.12/32 网段的 LDP LSP 一样。

LDP 跨域扩展仅需要在 Ingress 或 Transit 节点上进行配置，配置方法很简单，只需在 MPLS LDP 视图下执行 **longest-match** 命令，即可使能 LDP 按照最长匹配方式查找路由建立 LSP。缺省情况下，LDP 按照精确匹配方式查找路由建立 LSP。

4.2　LDP LSP 的 BFD 检测

在《华为路由器学习指南》一书中介绍了 BFD 在二、三层链路故障检测，以及与各种路由协议联动方面的应用。与 BFD 可以对二、三层链路进行快速的故障检测一样，BFD 也可以对 LSP 进行快速的故障检测，触发 LSP 在发生故障时进行快速主备路径倒换，提高整个网络的可靠性。在第 2 章已介绍了 BFD 在静态 LSP 检测中的应用及配置方法，本节要介绍 BFD 在 LDP 动态 LSP 检测方面的应用及配置方法。

4.2.1　BFD for LDP LSP

当采用 LDP LSP 承载流量时，如果主 LDP LSP 路径上的节点或链路发生故障时，如果有备份的 LDP LSP，则流量会向备份 LSP 切换。切换的速度依赖于故障的检测速度以及流量的切换速度，如果切换的速度很慢，将会导致长时间的流量丢失。其中流量的切换速度可以由 LDP FRR（Fast Reroute，快速重路由，将在 4.4 节介绍）来保证，但是由于 LDP 协议自身的故障检测机制检测速度较慢，所以仅仅采用 LDP FRR 技术并不能完全解决上述问题。

如图 4-2 所示的是一个存在 LDP LSP 主、备路径的网络示例，各 LSR 通过周期性地发送 Hello 消息，向邻居 LSR 通告它在网络中的存在，并维持 Hello 邻接关系。LSR 为每个邻居建立一个 Hello 保持定时器，用于维护 Hello 邻接关系，每收到一个 Hello 消息时刷新 Hello 保持定时器。如果在收到新的 Hello 消息之前 Hello 保持定时器超时，则 LSR 认为 Hello 邻接关系中断。这种机制并不能快速感知到网络的链路故障，尤其是 LSR 之间存在二层设备时。

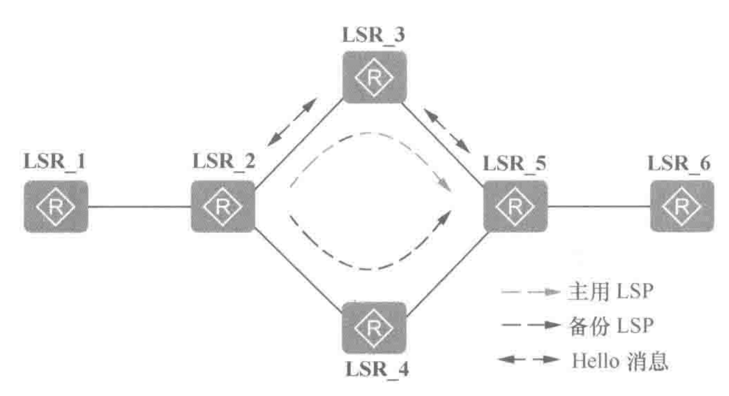


图 4-2 LDP LSP 主备路径

引入了 BFD 这种快速检测机制，可以对 LDP LSP 进行快速的故障检测，触发流量快速向备份路径切换，使得流量丢失最少，进一步提高业务的可靠性。

BFD for LDP LSP 是对 LDP LSP 的检测，能够快速检测到 LSP 的故障，并及时通知转发层面，从而保证流量的快速切换。

通过将 BFD 会话与 LSP 绑定，即可在入节点和出节点之间建立 BFD 会话。BFD 报文从源端开始经过 LSP 转发到达目的端，目的端再对该 BFD 报文进行回应，通过此方式在源端可以快速检测出 LSP 的状态。当检测出 LSP 故障以后，BFD 将此信息上报给设备转发层，然后设备转发层查找备份 LSP，将业务流量切换到备份 LSP 上。

BFD for LDP LSP 中可以采用静态 BFD，也可采用动态 BFD，下面分别予以介绍。

4.2.2 配置静态 BFD 检测 LDP LSP

通过配置静态 BFD 检测 LDP LSP，实现快速检测 LDP LSP 链路的目的。这种方式需人为控制，部署比较灵活。但在部署静态 BFD 检测 LDP LSP 时，需注意以下事项。

- 只能在 LDP LSP 的 Ingress 节点上进行 BFD 绑定。
- 一条 LSP 只能与一个 BFD 会话绑定。
- **只支持 32 位掩码的主机路由触发建立的 LDP LSP**，不支持其他路由触发建立的 LDP LSP。
- 往/返转发方式可以不一致（如报文从源端到目的端使用 LSP 转发，从目的端到源端使用 IP 转发），但要求往返路径一致，如果不一致，则检测到故障时，不能确定具体是哪条路径的故障。

另外，在配置静态 BFD 检测 LDP LSP 之前要配置好骨干网各节点，包括备份 LSP 途经的各节点间的本地 LDP 会话。

在配置静态 BFD 检测 LDP LSP 的过程中要先配置入节点 BFD 参数，然后配置出节点 BFD 参数，下面分别予以介绍。

1. 配置入节点 BFD 参数

入节点可配置的 BFD 参数包括：所绑定的本地静态 LSP、本地标识符、远端标识符、本地发送 BFD 报文的时间间隔、本地接收 BFD 报文的时间间隔和本地 BFD 检测倍数，这些将会影响会话的建立。用户可以根据网络的实际状况调整本地检测时间。对于不太

稳定的链路，如果本地检测时间较小，则 BFD 会话可能会发生震荡，这时可以选择增大本地检测时间。入节点的 BFD 参数配置步骤见表 4-1。

表 4-1　配置入节点 BFD 参数的步骤

步骤	命令	说明
1	**system-view** 例如：<Huawei> **system-view**	进入系统视图
2	**bfd** 例如：[Huawei] **bfd**	对本节点使能全局 BFD 能力并进入 BFD 全局视图。 缺省情况下，全局 BFD 功能未使能，可用 **undo bfd** 命令全局去使能 BFD 功能，如果已经配置了 BFD 会话信息，则所有的 BFD 会话都会被删除
3	**quit** 例如：[Huawei-bfd] **quit**	返回系统视图
4	**bfd** *cfg-name* **bind ldp-lsp peer-ip** *ip-address* **nexthop** *ip-address* [**interface** *interface-type interface-number*] 例如：[Huawei] **[Huawei] bfd** 1to4 **bind ldp-lsp peer-ip** 4.4.4.4 **nexthop** 1.1.1.1 **interface** gigabitethernet 1/0/0	配置 BFD 会话所绑定的静态 LSP。命令中的参数说明如下。 • *cfg-name*：指定 BFD 配置名，字符串形式，不支持空格，不区分大小写，长度范围是 1～15。当输入的字符串两端使用双引号时，可在字符串中输入空格。 • **peer-ip** *ip-address* ：指定 BFD 会话绑定动态 LSP 的目的端 IP 地址，必须是 MPLS LSR ID 或 LDP 实例的 LSR ID。 • **nexthop** *ip-address*：指定被检测 LSP 的下一跳 IP 地址。 • **interface** *interface-type interface-number*：可选参数，指定 BFD 绑定的出接口。当被检测的 LSP 出接口地址是借用的或者是被借用时，必须指定出接口。 缺省情况下，没有创建检测 LDP LSP 的 BFD 会话，可用 **undo bfd** *cfg-name* 命令删除指定的 BFD 会话。当 LDP LSP 被删除，但 LDP 会话没有删除时，与之绑定的 BFD 会话不会被删除，只是状态变为 down
5	**discriminator local** *discr-value* 例如：[Huawei-bfd-session-1to4] **discriminator local** 10	配置本地标识符，整数形式，取值范围是 1～8191。 BFD 会话两端设备的本地标识符和远端标识符需要分别对应，即本端的本地标识符与对端的远端标识符相同，否则会话无法正确建立。并且，本地标识符和远端标识符配置成功后不可修改，如果需要修改静态 BFD 会话本地标识符或者远端标识符，则必须先删除该 BFD 会话，然后再配置本地标识符
6	**discriminator remote** *discr-value* 例如：[Huawei-bfd-session-1to4] **discriminator remote** 20	配置远端标识符，整数形式，取值范围是 1～8191
7	**min-tx-interval** *interval* 例如：[Huawei-bfd-session-1to4] **min-tx-interval** 300	（可选）调整本地发送 BFD 报文的时间间隔，整数形式，取值范围是 10～2000，单位是 ms。 如果 BFD 会话在设置的检测周期内没有收到对端发来的 BFD 报文，则认为链路发生了故障，BFD 会话的状态将会置为 down。为降低对系统资源的占用，一旦检测到 BFD 会话状态变为 down，系统自动将本端的发送间隔调整为大于 1000ms 的一个随机值，当 BFD 会话的状态重新变为 Up 后，再恢复成用户配置的时间间隔

（续表）

步骤	命令	说明
7	**min-tx-interval** *interval* 例如：[Huawei-bfd-session-1to4] **min-tx-interval** 300	【说明】用户可以根据网络的实际状况增大或者降低 BFD 报文的发送和接收时间间隔。BFD 报文的发送、接收时间间隔直接决定了 BFD 会话的检测时间。对于不太稳定的链路，如果配置的 BFD 报文发送、接收时间间隔较小，则 BFD 会话可能会发生震荡，这时可以选择增大 BFD 报文的发送和接收时间间隔。通常情况下，建议使用缺省值。 缺省情况下，发送间隔是 1000ms，可用 **undo min-tx-interval** 命令恢复 BFD 报文的发送间隔为缺省值
8	**min-rx-interval** *interval* 例如：[Huawei-bfd-session-1to4] **min-rx-interval** 600	（可选）调整本地接收 BFD 报文的时间间隔，整数形式，取值范围是 10～2000，单位是 ms。 缺省情况下，接收间隔是 1000ms，可用 **undo min-rx-interval** 命令恢复 BFD 报文的接收间隔为缺省值
9	**detect-multiplier** *multiplier* 例如：[Huawei-bfd-session-1to4] **detect-multiplier** 5	（可选）调整本地 BFD 检测倍数，整数形式，取值范围是 3～50 BFD 会话的本端检测倍数直接决定了对端 BFD 会话的检测时间，检测时间 = 接收到的远端 Detect Multi × max（本地的 RMRI，接收到的 DMTI），其中，Detect Mult 是检测倍数，通过本条命令配置；RMRI 是本端能够支持的最短 BFD 报文接收间隔；DMTI 是本端想要采用的最短 BFD 报文的发送间隔。 【说明】用户可以根据网络的实际状况增大或者降低 BFD 会话的本地检测倍数。比如对于比较稳定的链路，由于不需要频繁的检测链路状态，因此可以增大 BFD 会话的检测倍数。 缺省情况下，本地 BFD 检测倍数为 3，可用 **undo detect-multiplier** 命令恢复 BFD 会话的本地检测倍数为缺省值
10	**process-pst** 例如：[Huawei-bfd-session-1to4] **process-pst**	允许 BFD 会话状态改变时通告上层应用。如果允许 BFD 修改端口状态表 PST（Port State Table），当检测到 BFD 会话状态变为 down 时，系统将更改 PST 中相应表项。 缺省情况下，静态 BFD 会话未使能通告联动检测业务，可用 **undo process-pst** 命令恢复缺省配置
11	**commit** 例如：[Huawei-bfd-session-1to4] **commit**	提交配置。无论改变任何 BFD 配置，必须执行 commit 命令，才能使配置生效。 【说明】BFD 会话建立需要满足一定的条件，包括绑定的接口状态是 Up、有去往 **peer-ip** 的可达路由。如果当前不满足会话建立条件，执行本命令后，系统将保留该会话的配置表项，但会话表项不能建立。但系统定期扫描已经提交但尚未建立会话的 BFD 配置表项，如果满足条件，则建立会话。 系统所允许建立的 BFD 会话有数量限制。当已经建立的 BFD 会话数达到上限时，如果对新的 BFD 会话执行本命令，系统将产生日志信息，提示无法创建会话，同时发送 Trap 消息

2. 配置出节点 BFD 参数

出节点可配置的 BFD 参数包括：本地标识符、远端标识符、本地发送 BFD 报文的时间间隔、本地接收 BFD 报文的时间间隔和本地 BFD 检测倍数，这些将会影响

BFD 会话的建立。用户可以根据网络的实际状况调整本地检测时间。对于不太稳定的链路，如果本地检测时间较小，则 BFD 会话可能会发生震荡，这时可以选择增大本地检测时间。

出节点的 BFD 参数配置步骤见表 4-2，与入节点的 BFD 会话配置方法基本一样，只不过在创建 BFD 会话时可根据反向通道的不同类型，要选择不同的配置命令。为了保证 BFD 报文往返路径一致，一般情况下反向通道优先选用 LSP 或者 TE 隧道。

表 4-2　配置出节点 **BFD** 参数的步骤

步骤	命令	说明
1	**system-view** 例如：<Huawei> **system-view**	进入系统视图
2	**bfd** 例如：[Huawei] **bfd**	对本节点使能全局 BFD 能力并进入 BFD 全局视图。 缺省情况下，全局 BFD 功能未使能，可用 **undo bfd** 命令全局去使能 BFD 功能，如果已经配置了 BFD 会话信息，则所有的 BFD 会话都会被删除
3	**quit** 例如：[Huawei-bfd] **quit**	返回系统视图
4	**bfd** *cfg-name* **bind peer-ip** *peer-ip* [**vpn-instance** *vpn-instance-name*] [**interface** *interface-type interface-number*] [**source-ip** *source-ip*] 例如：[Huawei] **bfd** 1to4 **bind peer-ip** 10.10.20.2	（四选一）当反向通道是 IP 链路时创建 BFD 会话。在创建 BFD 会话时，单跳检测必须绑定对端 IP 地址和本端相应接口，多跳检测只需绑定对端 IP 地址。命令中的参数说明如下。 • *cfg-name*：指定 BFD 配置名，字符串形式，不支持空格，不区分大小写，长度范围是 1～15。当输入的字符串两端使用双引号时，可在字符串中输入空格。 • **peer-ip** *ip-address*：指定 BFD 会话绑定的对端 IP 地址。**如果只指定对端 IP 地址，则表示检测多跳链路。** • **vpn-instance** *vpn-name*：可选参数，指定对端 BFD 会话绑定的 VPN 实例名称，必须是已创建的 VPN 实例。如果不指定 VPN 实例，则认为对端地址是公网地址。如果同时指定了对端 IP 地址和 VPN 实例，则表示检测 VPN 路由的多跳链路。 • **interface** *interface-type interface-number*：可选参数，指定绑定 BFD 会话的接口。**如果同时指定了对端 IP 地址和本端接口，表示检测单跳链路**，即检测以该接口为出接口、以 peer-ip 为下一跳地址的一条固定路由；如果同时指定了对端 IP 地址、VPN 实例和本端接口，表示检测 VPN 路由的单跳链路。 • **source-ip** *ip-address*：可选参数，指定 BFD 报文携带的源 IP 地址。通常情况下，不需要配置该参数。在 BFD 会话协商阶段，如果不配置该参数，则系统将在本地路由表中查找去往对端 IP 地址的出接口，以该出接口的 IP 地址作为本端发送 BFD 报文的源 IP 地址；在 BFD 会话检测链路阶段，如果不配置该参数，则系统会将 BFD 报文的源 IP 地址设置为一个固定的值 缺省情况下，没有创建 BFD 会话，可用 **undo bfd session-name** 命令删除指定的 BFD 会话，同时取消 BFD 会话的绑定信息

（续表）

步骤	命令	说明
4	**bfd** *cfg-name* **bind static-lsp** *lsp-name* 例如：[Huawei] **bfd** 1to4 **bind static-lsp** 1to4	（四选一）当反向通道是静态 LSP 时创建静态 LSP 的 BFD 会话。参数 *cfg-name* 用来指定所创建的 BFD 会话名称，*lsp-name* 指定 BFD 会话所绑定的静态 LSP 名称
	bfd *cfg-name* **bind ldp-lsp peer-ip** *ip-address* **nexthop** *ip-address* [**interface** *interface-type interface-number*] 例如：Huawei] **bfd** 1to4 **bind ldp-lsp peer-ip** 4.4.4.4 **nexthop** 1.1.1.1 **interface** gigabitethernet 1/0/0	（四选一）当反向通道是动态 LSP 时创建 LDP LSP 的 BFD 会话。命令中的参数说明如下。 • *cfg-name*：指定 BFD 会话名称，字符串形式，不支持空格，不区分大小写，长度范围是 1～15。当输入的字符串两端使用双引号时，可在字符串中输入空格。 • **peer-ip** *ip-address*：指定 BFD 会话绑定动态 LSP 的目的端 IP 地址。 • **nexthop** *ip-address*：指定被检测 LSP 的下一跳 IP 地址。 • **interface** *interface-type interface-number* ：可选参数，指定 BFD 绑定的出接口。 缺省情况下，没有创建检测 LDP LSP 的 BFD 会话，可用 **undo bfd** *cfg-name* 命令删除指定的 BFD 会话
	bfd *cfg-name* **bind mpls-te interface tunnel** *interface-number* [**te-lsp** [**backup**]] 例如：[Huawei] **bfd** 1to4 **bind mpls-te interface** Tunnel 0/0/1 **te-lsp**	（四选一）当反向通道是 TE 隧道时创建 BFD 会话或与 TE 隧道绑定的主用或备用 LSP。命令中的参数说明如下。 • *cfg-name*：指定创建 BFD 会话名称，字符串形式，不支持空格，不区分大小写，长度范围是 1～15。当输入的字符串两端使用双引号时，可在字符串中输入空格。 • **interface tunnel** *interface-number*：指定 BFD 会话绑定的 Tunnel 接口编号。 • **te-lsp** [**backup**]：可选项，指定 BFD 检测与 Tunnel 隧道绑定的 LSP。其中：未选择 **backup** 可选项时，指定 BFD 检测与 Tunnel 隧道绑定的主 LSP；选择了 **backup** 可选项时，指定 BFD 检测与 Tunnel 隧道绑定的备用 LSP。BFD 检测与 Tunnel 绑定的主用或备用 LSP 时，如果 LSP 的状态为 down，则不能建立 BFD 会话。 BFD 检测 TE 隧道时，如果 TE 隧道的状态为 Down，则能够创建 BFD 会话，但 BFD 会话不能 Up。一个 TE 隧道可能有多个 LSP，当 BFD 检测 TE 隧道时，只有全部 LSP 都出现故障时，BFD 会话的状态才为 down。 缺省情况下，Tunnel 隧道没有使用 BFD 检测，可用 **undo bfd** *cfg-name* 命令删除指定的 BFD 会话
5	**discriminator local** *discr-value* 例如：[Huawei-bfd-session-1to4] **discriminator** local 10	配置本地标识符，参见表 4-1 中的第 5 步
6	**discriminator remote** *discr-value* 例如：[Huawei-bfd-session-1to4] **discriminator** remote 20	配置远端标识符，参见表 4-1 中的第 6 步
7	**min-tx-interval** *interval* 例如：[Huawei-bfd-session-1to4] **min-tx-interval** 300	（可选）调整本地发送 BFD 报文的时间间隔，参见表 4-1 中的第 7 步

（续表）

步骤	命令	说明
8	**min-rx-interval** *interval* 例如：[Huawei-bfd-session-1to4] **min-rx-interval** 600	（可选）调整本地接收 BFD 报文的时间间隔，参见表 4-1 中的第 8 步
9	**detect-multiplier** *multiplier* 例如：[Huawei-bfd-session-1to4] **detect-multiplier** 5	（可选）调整本地 BFD 检测倍数，参见表 4-1 中的第 9 步
10	**process-pst** 例如：[Huawei-bfd-session-1to4] **process-pst**	（可选）允许 BFD 会话状态改变时通告上层应用，参见表 4-1 中的第 10 步
11	**commit** 例如：[Huawei-bfd-session-1to4] **commit**	提交配置，参见表 4-1 中的第 11 步

4.2.3　配置动态 BFD 检测 LDP LSP

配置动态 BFD 检测 LDP LSP，不需要指定 BFD 参数，能够提高链路故障检测速度、减少配置工作量。这种方式配置简单，更具灵活性。

动态 BFD 检测 LDP LSP 时，需注意以下两点。

- 只支持主机路由触发建立的 LDP LSP。
- 往返转发方式可以不一致（如报文从源端到目的端使用 LSP 转发，从目的端到源端使用 IP 转发），但要求往返路径一致，如果不一致，则检测到故障时，不能确定具体是哪条路径的故障。

在配置动态 BFD 检测 LDP LSP 之前也需要配置好本地 LDP 会话。然后按照以下顺序进行配置。

（1）使能全局 BFD 能力。

（2）使能 MPLS 动态创建 BFD 会话功能。

（3）配置动态 BFD 检测 LDP LSP 的触发策略。

（4）（可选）调整 BFD 检测参数。

下面对以上配置任务的具体配置方法分别予以介绍。

1. 使能全局 BFD 能力

只有全局使能 BFD 功能后，才能进行 BFD 的相关配置。需要在源端和目的端分别配置，配置方法很简单，只需在系统视图下执行 **bfd** 命令即可。缺省情况下，全局 BFD 功能未使能，可用 **undo bfd** 命令全局去使能 BFD 功能，此时 BFD 的所有功能将会关闭。如果已经配置了 BFD 会话信息，则所有的 BFD 会话都会被删除。

2. 使能 MPLS 动态创建 BFD 会话功能

在源端和目的端上使能 BFD 能力，就可以使能 MPLS 动态创建 BFD 会话功能。

在源端的配置方法是在 MPLS 视图下执行 **mpls bfd enable** 命令，使能 LDP LSP 动态创建 BFD 会话的能力。缺省情况下，在 LDP LSP 的源端设备上禁止动态创建 BFD 会

话的功能。但执行完该命令并不会立即创建 BFD 会话。

在目的端的配置方法是在 BFD 视图下执行 **mpls-passive** 命令，使能被动创建 BFD 会话功能。缺省情况下，不使能被动动态创建 BFD 会话功能。执行完该命令也不会立即创建 BFD 会话，而是等接收到源端发送的携带 BFD TLV 的 LSP ping 请求报文后才会触发建立 BFD 会话。

3. 配置动态 BFD 检测 LDP LSP 的触发策略

动态 BFD 检测 LDP LSP 的触发策略有两种。

■ 主机触发：如果需要所有的主机地址均能触发 BFD 会话的建立，则采用主机触发方式。还可以通过指定 nexthop（下一跳）和 outgoing-interface（出接口）来约束哪些 LSP 可以建立 BFD 会话。

■ FEC 列表触发：如果只需要其中的一部分主机触发 BFD 会话的建立，可以采用 fec-list 触发方式，来指定相应的主机地址。

可以根据需要在被检测 LSP 的**源端**上进行配置，具体的配置方法如表 4-3 所示。

表 4-3 配置动态 **BFD** 检测 **LDP LSP** 触发策略的步骤

步骤	命令	说明
1	**system-view** 例如：<Huawei> **system-view**	进入系统视图
2	**fec-list** *list-name* 例如：[Huawei] **fec-list** feclist	（可选）创建 FEC 列表，进入该列表视图。仅当采用 FEC 列表触发 BFD 会话时才需要配置，全局只能创建一个 FEC 列表。 参数 *list-name* 用来指定 FEC 列表的名称，字符串形式，不支持空格，**区分大小写**，长度范围是 1～31。当输入的字符串两端使用双引号时，可在字符串中输入空格。 缺省情况下，没有创建 FEC 列表，可用 **undo fec-list** *list-name* 命令删除指定的 FEC 列表
3	**fec-node** *ip-address* [**nexthop** *ip-address* | **outgoing-interface** [*interface-type interface-number*]]* 例如：[Huawei-fec-list-feclist] **fec-node** 2.2.2.2 **nexthop** 100.1.2.1 **outgoing-interface** gigabitethernet 1/0/0	（可选）在当前 FEC 列表中增加 FEC 节点，指定主机路由触发建立 BFD 会话，可多次配置本命令，添加多个 FEC 节点。仅当采用 FEC 列表触发 BFD 会话时才需要配置。命令中的参数说明如下。 • *ip-address*：指定 FEC 的 IP 地址。 • **nexthop** *ip-address*：可多选可选参数，指定下一跳地址。 • **outgoing-interface** *interface-type interface-number*：可多选可选参数，指定出接口。如果不指定具体的接口，则可是任意出接口。 缺省情况下，没有增加 FEC 节点，可用 **undo fec-node** *ip-address* [**nexthop** *ip-address* | **outgoing-interface** [*interface-type interface-number*]]*命令删除指定的 FEC 节点
4	**quit** 例如：[Huawei-fec-list-feclist] **quit**	返回系统视图
5	**mpls** 例如：[Huawei] **mpls**	进入 MPLS 视图

（续表）

步骤	命令	说明
6	**mpls bfd-trigger** [**host** [**nexthop** *next-hop-address* | **outgoing-interface** *interface-type interface-number*] * | **fec-list** *list-name*] 例如：[Huawei-mpls] **mpls bfd-trigger host**	配置动态 BFD 检测 LDP LSP 的触发策略，执行完该命令才真正开始创建 BFD 会话。命令中的参数和选项说明如下。 ● **host**：二选一可选项，指定 LDP BFD 以所有主机方式触发。 ● **nexthop** *next-hop-address*：可多选可选参数，指定 LSP 的下一跳地址。 ● **outgoing-interface** *interface-type interface-number*：可多选可选参数，指定 LSP 的出接口。 ● **fec-list** *list-name*:二选一可选参数，指定 LDP BFD 以 FEC 列表方式触发，并指定 FEC 列表的名称。 如果以上参数和选项都不配置，则采用的是主机触发方式，且所有的主机地址均能触发 BFD 会话的建立。 缺省情况下，没有配置 LDP BFD 触发策略，可用 **undo mpls bfd-trigger** [**host** [**nexthop** *next-hop-address* | **outgoing-interface** *interface-type interface-number*] * | fec-list *list-name*] 命令删除指定的 LDP BFD 触发策略

4.（可选）调整 BFD 检测参数

配置 BFD 检测参数包括本地发送 BFD 报文的时间间隔、本地接收 BFD 报文的时间间隔和本地 BFD 检测倍数，这些将会影响会话的建立。

在动态 BFD 检测动态 LDP LSP 的配置中，检测参数的具体配置步骤如表 4-4 所示。用户可以根据网络的实际状况调整本地检测时间。对于不太稳定的链路，如果本地检测时间较小，则 BFD 会话可能会发生震荡，这时可以选择增大 BFD 检测的参数。

本地实际发送 BFD 报文的时间间隔＝MAX｛本地配置发送 BFD 报文时间间隔，对端配置接收 BFD 报文时间间隔｝；本地实际接收 BFD 报文时间间隔＝MAX｛对端配置发送 BFD 报文时间间隔，本地配置接收 BFD 报文时间间隔｝；本地检测时间＝本地实际接收 BFD 报文时间间隔×对端配置 BFD 检测倍数。

表 4-4　　调整动态 BFD 检测 LDP LSP BFD 检测参数的配置步骤

步骤	命令	说明
1	**system-view** 例如：<Huawei> **system-view**	进入系统视图
2	**bfd** 例如：[Huawei] **bfd**	进入 BFD 视图
3	**mpls ping interval** *interval* 例如：[Huawei-bfd] **mpls ping interval** 100	调节发送 LSP ping 报文的时间间隔，整数形式，取值范围是 30～600，缺省值是 60s。 缺省情况下，动态 BFD 中 LSP ping 定时器的时间间隔是 60s，可用 **undo mpls ping interval** 命令恢复动态 BFD 中 LSP ping 定时器的时间间隔为缺省值
4	**quit** 例如：[Huawei-bfd] **quit**	返回系统视图
5	**mpls** 例如：[Huawei] **mpls**	进入 MPLS 视图

（续表）

步骤	命令	说明
6	**mpls bfd** { **min-tx-interval** *interval* | **min-rx-interval** *interval* | **detect-multiplier** *multiplier* }* 例如：[Huawei-mpls] **mpls bfd min-tx-interval** 200	设置 BFD 检测的参数。命令中的参数说明如下。 • **min-tx-interval** *interval*：可多选参数，指定 BFD 会话发送时间间隔，整数形式，取值范围是 10～2000，单位是 ms，缺省值是 1000ms。 • **min-rx-interval** *interval*：可多选参数，指定 BFD 会话接收时间间隔，整数形式，取值范围是 10～2000，单位是 ms，缺省值是 1000ms。 • **detect-multiplier** *multiplier*：可多选参数，指定 BFD 会话本地检测的倍数，整数形式，取值范围是 3～50，缺省值是 3。 缺省情况下，没有设置BFD会话的相关参数，可用 **undo mpls bfd** { **min-tx-interval** | **min-rx-interval** | **detect-multiplier** }* 命令删除 BFD 会话的相关参数配置

4.2.4 BFD 检测 LDP LSP 维护和管理命令

在进行 BFD 检测 LDP LSP 配置或应用时，可使用以下命令进行维护或管理。

■ **display bfd configuration** { **all** | **static** }：查看所有或静态 BFD 会话配置信息。

■ **display bfd session** { **all** | **static** }：查看所有或静态 BFD 会话信息。

■ **display bfd statistics session** { **all** | **static** }：查看所有或静态 BFD 会话统计信息。

■ **display bfd configuration all** [**verbose**]：查看源端所有 BFD 会话配置信息。

■ **display bfd configuration passive-dynamic** [**peer-ip** *peer-ip* **remote-discriminator** *discriminator*] [**verbose**]：查看目的端所有或指定 BFD 会话配置信息。

■ **display bfd session all** [**verbose**]：查看源端所有 BFD 会话信息。

■ **display bfd session passive-dynamic** [**peer-ip** *peer-ip* **remote-discriminator** *discriminator*] [**verbose**]：查看目的端被动创建的所有或指定 BFD 会话信息。

■ **display mpls bfd session** [**statistics** | **protocol ldp** | **outgoing-interface** *interface-type interface-number* | **nexthop** *ip-address* | **fec** *fec-address* | **verbose** | **monitor**]：查看源端所有或指定 MPLS 的 BFD 会话信息。

4.2.5 静态 BFD 检测 LDP LSP 配置示例

如图 4-3 所示，网络拓扑结构简单并且稳定，在 PE1→P1→PE2 上建立 LDP LSP，PE2→P2→PE1 为 IP 链路。如果采用接口自己感知故障，则所花费的时间比较长。要求对 LDP LSP 进行连通性检测，当 LDP LSP 出现故障时，PE1 能够在 500ms 之内收到故障通告。

1. 基本配置思路分析

本示例的拓扑结构简单且稳定，可以通过配置静态 BFD 检测 LDP LSP 实现本示例的需求。需要在 PE1、PE2 上配置针对 LDP LSP 的静态 BFD 会话，具体的配置思路如下。

（1）在各 PE、P 上配置各接口（包括 Loopback 接口）的 IP 地址。

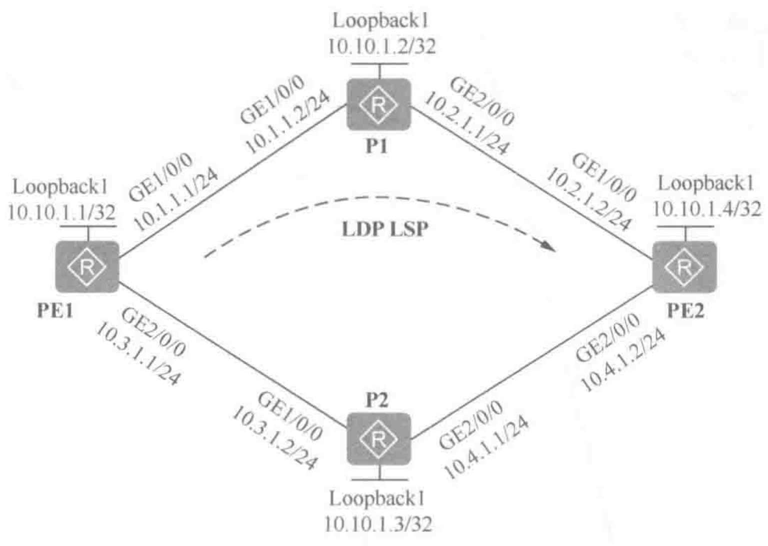


图 4-3　静态 BFD 检测 LDP LSP 配置示例的拓扑结构

（2）在各 PE、P 上配置 OSPF 路由，实现骨干网的 IP 连通性。

（3）在 PE1→P1→PE2 链路上配置 LDP 功能，建立 LDP LSP。

（4）在 PE1 和 PE2 之间配置静态 BFD 会话，以 PE1 为入节点，PE2 为出节点。调整 BFD 检测参数，实现 PE1 能够在 500ms 之内收到故障通告。

2. 具体配置步骤

（1）配置各接口的 IP 地址。

PE1 上的配置。

```
<Huawei> system-view
[Huawei] sysname PE1
[PE1] interface loopback 1
[PE1-LoopBack1] ip address 10.10.1.1 32
[PE1-LoopBack1] quit
[PE1] interface gigabitethernet 1/0/0
[PE1-GigabitEthernet1/0/0] ip address 10.1.1.1 24
[PE1-GigabitEthernet1/0/0] quit
[PE1] interface gigabitethernet 2/0/0
[PE1-GigabitEthernet2/0/0] ip address 10.3.1.1 24
[PE1-GigabitEthernet2/0/0] quit
```

P1 上的配置。

```
<Huawei> system-view
[Huawei] sysname P1
[P1] interface loopback 1
[P1-LoopBack1] ip address 10.10.1.2 32
[P1-LoopBack1] quit
[P1] interface gigabitethernet 1/0/0
[P1-GigabitEthernet1/0/0] ip address 10.1.1.2 24
[P1-GigabitEthernet1/0/0] quit
[P1] interface gigabitethernet 2/0/0
[P1-GigabitEthernet2/0/0] ip address 10.2.1.1 24
[P1-GigabitEthernet2/0/0] quit
```

P2 上的配置。

```
<Huawei> system-view
[Huawei] sysname P2
```

```
[P2] interface loopback 1
[P2-LoopBack1] ip address 10.10.1.3 32
[P2-LoopBack1] quit
[P2] interface gigabitethernet 1/0/0
[P2-GigabitEthernet1/0/0] ip address 10.3.1.2 24
[P2-GigabitEthernet1/0/0] quit
[P2] interface gigabitethernet 2/0/0
[P2-GigabitEthernet2/0/0] ip address 10.4.1.1 24
[P2-GigabitEthernet2/0/0] quit
```

PE2 上的配置。

```
<Huawei> system-view
[Huawei] sysname PE2
[PE2] interface loopback 1
[PE2-LoopBack1] ip address 10.10.1.4 32
[PE2-LoopBack1] quit
[PE2] interface gigabitethernet 1/0/0
[PE2-GigabitEthernet1/0/0] ip address 10.2.1.2 24
[PE2-GigabitEthernet1/0/0] quit
[PE2] interface gigabitethernet 2/0/0
[PE2-GigabitEthernet2/0/0] ip address 10.4.1.2 24
[PE2-GigabitEthernet2/0/0] quit
```

（2）配置 OSPF 协议发布各节点接口所连网段和 LSR ID 的主机路由，都加入到缺省的 OSPF 1 进程，区域 0 中。

PE1 上的配置。

```
[PE1] ospf 1
[PE1-ospf-1] area 0
[PE1-ospf-1-area-0.0.0.0] network 10.10.1.1 0.0.0.0
[PE1-ospf-1-area-0.0.0.0] network 10.1.1.0 0.0.0.255
[PE1-ospf-1-area-0.0.0.0] network 10.3.1.0 0.0.0.255
[PE1-ospf-1-area-0.0.0.0] quit
[PE1-ospf-1] quit
```

P1 上的配置。

```
[P1] ospf 1
[P1-ospf-1] area 0
[P1-ospf-1-area-0.0.0.0] network 10.10.1.2 0.0.0.0
[P1-ospf-1-area-0.0.0.0] network 10.1.1.0 0.0.0.255
[P1-ospf-1-area-0.0.0.0] network 10.2.1.0 0.0.0.255
[P1-ospf-1-area-0.0.0.0] quit
[P1-ospf-1] quit
```

P2 上的配置。

```
[P2] ospf 1
[P2-ospf-1] area 0
[P2-ospf-1-area-0.0.0.0] network 10.10.1.3 0.0.0.0
[P2-ospf-1-area-0.0.0.0] network 10.3.1.0 0.0.0.255
[P2-ospf-1-area-0.0.0.0] network 10.4.1.0 0.0.0.255
[P2-ospf-1-area-0.0.0.0] quit
[P2-ospf-1] quit
```

PE2 上的配置。

```
[PE2] ospf 1
[PE2-ospf-1] area 0
[PE2-ospf-1-area-0.0.0.0] network 10.10.1.4 0.0.0.0
[PE2-ospf-1-area-0.0.0.0] network 10.2.1.0 0.0.0.255
```

```
[PE2-ospf-1-area-0.0.0.0] network 10.4.1.0 0.0.0.255
[PE2-ospf-1-area-0.0.0.0] quit
[PE2-ospf-1] quit
```

（3）在 PE1→P1→PE2 上配置 LDP 功能，建立 LDP LSP。因为本示例明确指出 PE2→P1→PE1 为 IP 链路，即采用 IP 路由转发，故不需要在这条链路上配置 LDP LSP。

\# PE1 上的配置。

```
[PE1] mpls lsr-id 10.10.1.1
[PE1] mpls
[PE1-mpls] quit
[PE1] mpls ldp
[PE1-mpls-ldp] quit
[PE1] interface gigabitethernet 1/0/0
[PE1-GigabitEthernet1/0/0] mpls
[PE1-GigabitEthernet1/0/0] mpls ldp
[PE1-GigabitEthernet1/0/0] quit
```

\# P1 上的配置。

```
[P1] mpls lsr-id 10.10.1.2
[P1] mpls
[P1-mpls] quit
[P1] mpls ldp
[P1-mpls-ldp] quit
[P1] interface gigabitethernet 1/0/0
[P1-GigabitEthernet1/0/0] mpls
[P1-GigabitEthernet1/0/0] mpls ldp
[P1-GigabitEthernet1/0/0] quit
[P1] interface gigabitethernet 2/0/0
[P1-GigabitEthernet2/0/0] mpls
[P1-GigabitEthernet2/0/0] mpls ldp
[P1-GigabitEthernet2/0/0] quit
```

\# PE2 上的配置。

```
[PE2] mpls lsr-id 10.10.1.4
[PE2] mpls
[PE2-mpls] quit
[PE2] mpls ldp
[PE2-mpls-ldp] quit
[PE2] interface gigabitethernet 1/0/0
[PE2-GigabitEthernet1/0/0] mpls
[PE2-GigabitEthernet1/0/0] mpls ldp
[PE2-GigabitEthernet1/0/0] quit
```

\# 执行 **display mpls ldp lsp** 命令，可以看到在 PE1 上建立了到目的地址为 10.10.1.4/32 的 LDP LSP，参见输出信息中的粗体字部分。

```
[PE1] display mpls ldp lsp

 LDP LSP Information
 -------------------------------------------------------------------------------
 DestAddress/Mask   In/OutLabel    UpstreamPeer    NextHop      OutInterface
 -------------------------------------------------------------------------------
 10.10.1.1/32       3/NULL         10.10.1.2       127.0.0.1    InLoop0
*10.10.1.1/32       Liberal/3                      DS/10.10.1.2
 10.10.1.2/32       NULL/3         -               10.1.1.2     GE1/0/0
 10.10.1.2/32       1024/3         10.10.1.2       10.1.1.2     GE1/0/0
 10.10.1.4/32       NULL/1025      -               10.1.1.2     GE1/0/0
```

```
10.10.1.4/32        1022/1025        10.10.1.2        10.1.1.2     GE1/0/0
-------------------------------------------------------------------------------
TOTAL: 5 Normal LSP(s) Found.
TOTAL: 1 Liberal LSP(s) Found.
TOTAL: 0 Frr LSP(s) Found.
A '*' before an LSP means the LSP is not established
A '*' before a Label means the USCB or DSCB is stale
A '*' before a UpstreamPeer means the session is state
A '*' before a DS means the session is stale
A '*' before a NextHop means the LSP is FRR LSP
```

（4）配置 PE1 和 PE2 之间的静态 BFD 会话，调整发送报文的最小时间间隔是 100ms，接收报文的最小时间间隔是 100ms。

入节点 PE1 上的配置。

```
[PE1] bfd
[PE1-bfd] quit
[PE1] bfd pe1tope2 bind ldp-lsp peer-ip 10.10.1.4 nexthop 10.1.1.2 interface gigabitethernet 1/0/0
[PE1-bfd-lsp-session-pe1tope2] discriminator local 1   #--要与对端的远端标识符一致
[PE1-bfd-lsp-session-pe1tope2] discriminator remote 2   #--要与对端的本地标识符一致
[PE1-bfd-lsp-session-pe1tope2] min-tx-interval 100   #---调整 BFD 检测报文的发送时间间隔为 100 毫秒
[PE1-bfd-lsp-session-pe1tope2] min-rx-interval 100   #---调整 BFD 检测报文的接收时间间隔为 100 毫秒
[PE1-bfd-lsp-session-pe1tope2] process-pst
[PE1-bfd-lsp-session-pe1tope2] commit
[PE1-bfd-lsp-session-pe1tope2] quit
```

出节点 PE2 上的配置。

```
[PE2] bfd
[PE2-bfd] quit
[PE2] bfd pe2tope1 bind  peer-ip 10.10.1.1
[PE2-bfd-session-pe2tope1] discriminator local 2
[PE2-bfd-session-pe2tope1] discriminator remote 1
[PE2-bfd-session-pe2tope1] min-tx-interval 100
[PE2-bfd-session-pe2tope1] min-rx-interval 100
[PE2-bfd-session-pe2tope1] commit
[PE2-bfd-session-pe2tope1] quit
```

3. 实验结果验证

以上配置完成后，在 Ingress 上执行 **display bfd session all** 命令，可以看到它与出节点 PE2 之间建立的 BFD 会话状态（State 字段）的值为“Up”，表示已成功建立了 BFD 会话。

```
[PE1] display bfd session all
-------------------------------------------------------------------------------
Local Remote     PeerIpAddr      State     Type        InterfaceName
-------------------------------------------------------------------------------
1     2          10.10.1.4       Up        S_LDP_LSP   GigabitEthernet1/0/0
-------------------------------------------------------------------------------
      Total Up/DOWN Session Number : 1/0
```

在 Egress 上执行 **display bfd session all** 命令，也可以看到它与入节点 PE1 之间建立的 BFD 会话状态为 Up。

```
[PE2] display bfd session all
-------------------------------------------------------------------------------
Local Remote     PeerIpAddr      State     Type        InterfaceName
-------------------------------------------------------------------------------
2     1          10.10.1.1       Up        S_IP_PEER   -
```

```
------------------------------------------------------------------------------
  Total Up/DOWN Session Number : 1/0
```

4.2.6 动态 BFD 检测 LDP LSP 配置示例

如图 4-4 所示，网络拓扑结构复杂并且不稳定，节点 LSRA、LSRB 和 LSRC 于同一 MPLS 域，LSRA 和 LSRC 间创建 LDP LSP 链路。如果采用接口自己感知故障，则所花费的时间比较长。现要求对 LDP LSP 进行连通性检测，当 LDP LSP 出现故障时，LSRA 能够在 500ms 之内收到故障通告。



图 4-4 动态 BFD 检测 LDP LSP 配置示例的拓扑结构

1. 基本配置思路分析

由于本示例的网络拓扑结构复杂并且不稳定，因此可在 LSRA、LSRC 上配置 BFD 会话，检测 LSRA 和 LSRC 之间的 LDP LSP。但在配置动态 BFD 检测 LDP LSP 前仍需要完成 LDP 必选基本功能的配置。故本示例的基本配置思路如下。

（1）在各 LSR 上配置各接口（包括 Loopback 接口）的 IP 地址。

（2）在各 LSR 上配置 OSPF 路由，实现骨干网的 IP 连通性。

（3）在各 LSR 上配置 LDP，使各 LSR 间可建立到达对方 LSR ID 所代表的主机路由的 LDP LSP。

（4）LSRA 和 LSRC 之间的动态 BFD 会话，配置调整 BFD 检测参数，实现 LSRA 能够在 500ms 之内收到故障通告。

2. 具体配置步骤

（1）配置各接口（包括 Loopback 接口）的 IP 地址。

\# LSRA 上的配置。

```
<Huawei> system-view
[Huawei] sysname LSRA
[LSRA] interface loopback 0
[LSRA-LoopBack0] ip address 10.10.1.1 32
[LSRA-LoopBack0] quit
[LSRA] interface gigabitethernet 1/0/0
[LSRA-GigabitEthernet1/0/0] ip address 10.1.1.1 24
[LSRA-GigabitEthernet1/0/0] quit
```

\# LSRB 上的配置。

```
<Huawei> system-view
[Huawei] sysname LSRB
[LSRB] interface loopback 0
```

```
[LSRB-LoopBack0] ip address 10.10.1.2 32
[LSRB-LoopBack0] quit
[LSRB] interface gigabitethernet 1/0/0
[LSRB-GigabitEthernet1/0/0] ip address 10.1.1.2 24
[LSRB-GigabitEthernet1/0/0] quit
[LSRB] interface gigabitethernet 2/0/0
[LSRB-GigabitEthernet2/0/0] ip address 10.2.1.1 24
[LSRB-GigabitEthernet2/0/0] quit
```

LSRC 上的配置。

```
<Huawei> system-view
[Huawei] sysname LSRC
[LSRC] interface loopback 0
[LSRC-LoopBack0] ip address 10.10.1.3 32
[LSRC-LoopBack0] quit
[LSRC] interface gigabitethernet 1/0/0
[LSRC-GigabitEthernet1/0/0] ip address 10.2.1.2 24
[LSRC-GigabitEthernet1/0/0] quit
```

（2）配置 OSPF 协议发布各节点接口所连网段和 LSR ID 的主机路由，都加入到缺省的 OSPF 1 进程，区域 0 中。

LSRA 上的配置。

```
[LSRA] ospf 1
[LSRA-ospf-1] area 0
[LSRA-ospf-1-area-0.0.0.0] network 10.10.1.1 0.0.0.0
[LSRA-ospf-1-area-0.0.0.0] network 10.1.1.0 0.0.0.255
[LSRA-ospf-1-area-0.0.0.0] quit
[LSRA-ospf-1] quit
```

LSRB 上的配置。

```
[LSRB] ospf 1
[LSRB-ospf-1] area 0
[LSRB-ospf-1-area-0.0.0.0] network 10.10.1.2 0.0.0.0
[LSRB-ospf-1-area-0.0.0.0] network 10.1.1.0 0.0.0.255
[LSRB-ospf-1-area-0.0.0.0] network 10.2.1.0 0.0.0.255
[LSRB-ospf-1-area-0.0.0.0] quit
[LSRB-ospf-1] quit
```

LSRC 上的配置。

```
[LSRC] ospf 1
[LSRC-ospf-1] area 0
[LSRC-ospf-1-area-0.0.0.0] network 10.10.1.3 0.0.0.0
[LSRC-ospf-1-area-0.0.0.0] network 10.2.1.0 0.0.0.255
[LSRC-ospf-1-area-0.0.0.0] quit
[LSRA-ospf-1] quit
```

（3）配置各 LSR 的 LDP 功能，建立各节点设备上主机路由对应的 LDP LSP。

LSRA 上的配置。

```
[LSRA] mpls lsr-id 10.10.1.1
[LSRA] mpls
[LSRA-mpls] quit
[LSRA] mpls ldp
[LSRA-ldp] quit
[LSRA] interface gigabitethernet 1/0/0
[LSRA-GigabitEthernet1/0/0] mpls
[LSRA-GigabitEthernet1/0/0] mpls ldp
[LSRA-GigabitEthernet1/0/0] quit
```

LSRB 上的配置。

```
[LSRB] mpls lsr-id 10.10.1.2
[LSRB] mpls
[LSRB-mpls] quit
[LSRB] mpls ldp
[LSRB-mpls-ldp] quit
[LSRB] interface gigabitethernet 1/0/0
[LSRB-GigabitEthernet1/0/0] mpls
[LSRB-GigabitEthernet1/0/0] mpls ldp
[LSRB-GigabitEthernet1/0/0] quit
[LSRB] interface gigabitethernet 2/0/0
[LSRB-GigabitEthernet2/0/0] mpls
[LSRB-GigabitEthernet2/0/0] mpls ldp
[LSRB-GigabitEthernet2/0/0] quit
```

LSRC 上的配置。

```
[LSRC] mpls lsr-id 10.10.1.3
[LSRC] mpls
[LSRC-mpls] quit
[LSRC] mpls ldp
[LSRC-ldp] quit
[LSRC] interface gigabitethernet 1/0/0
[LSRC-GigabitEthernet1/0/0] mpls
[LSRC-GigabitEthernet1/0/0] mpls ldp
[LSRC-GigabitEthernet1/0/0] quit
```

以上配置完成后，在 LSRA 执行 **display mpls ldp lsp** 命令，可以看到 LSRA 和 LSRC 之间的 LDP LSP 已经建立。以下是在 LSRA 上执行该命令的输出示例，从中可以看到它已有关于 LSRC 上 10.10.1.3/32 的 LDP LSP，参见输出信息粗体字部分。

```
[LSRA] display mpls ldp lsp

 LDP LSP Information
 -------------------------------------------------------------------------------
 DestAddress/Mask   In/OutLabel    UpstreamPeer    NextHop         OutInterface
 -------------------------------------------------------------------------------
 10.10.1.1/32       3/NULL         10.10.1.2       127.0.0.1       InLoop0
*10.10.1.1/32       Liberal/3                      DS/10.10.1.2
 10.10.1.2/32       NULL/3         -               10.1.1.2        GE1/0/0
 10.10.1.2/32       1024/3         10.10.1.2       10.1.1.2        GE1/0/0
 10.10.1.3/32       NULL/1025      -               10.1.1.2        GE1/0/0
 10.10.1.3/32       1025/1025      10.10.1.2       10.1.1.2        GE1/0/0
 -------------------------------------------------------------------------------
 TOTAL: 5 Normal LSP(s) Found.
 TOTAL: 1 Liberal LSP(s) Found.
 TOTAL: 0 Frr LSP(s) Found.
 A '*' before an LSP means the LSP is not established
 A '*' before a Label means the USCB or DSCB is stale
 A '*' before a UpstreamPeer means the session is stale
 A '*' before a DS means the session is stale
 A '*' before a NextHop means the LSP is FRR LSP
```

（4）配置动态 BFD 会话，检测 LSRA 和 LSRC 之间的 LDP LSP。

在 LSRA 配置 FEC 列表，这样就可以保障只触发 BFD 检测 LSRA 和 LSRC 之间的 LDP LSP。

```
[LSRA] fec-list tortc
```

```
[LSRA-fec-list-tortc] fec-node 10.10.1.3
[LSRA-fec-list-tortc] quit
```

在 LSRA 使能 BFD，指定动态触发 BFD 会话的 FEC 列表（即前面创建的 FEC 列表），并调整 BFD 检测参数，BFD 报文的发送和接收时间间隔均为 100ms。

```
[LSRA] bfd
[LSRA-bfd] quit
[LSRA] mpls
[LSRA-mpls] mpls bfd-trigger fec-list tortc
[LSRA-mpls] mpls bfd enable
[LSRA-mpls] mpls bfd min-tx-interval 100 min-rx-interval 100
[LSRA-mpls] quit
```

在 LSRC 上配置被动使能 BFD for LSP 能力。

```
[LSRC] bfd
[LSRC-bfd] mpls-passive
```

3. 配置结果验证。

以上配置完成后，在 LSRA 上执行 **display bfd session all** 命令，可查看前面与 LSRC 创建的动态 BFD 会话状态为 Up，表明动态 BFD 会话建立成功。

```
[LSRA] display bfd session all
--------------------------------------------------------------------------------
Local Remote     PeerIpAddr      State     Type         InterfaceName
--------------------------------------------------------------------------------
8192   8192        10.10.1.3        Up         D_LDP_LSP     GigabitEthernet1/0/0
--------------------------------------------------------------------------------
      Total Up/DOWN Session Number : 1/0
```

在 LSRC 上执行 **display bfd session all** 命令，也可查看到与 LSRA 创建的动态 BFD 会话状态为 Up。

```
[LSRC] display bfd session passive-dynamic
--------------------------------------------------------------------------------
Local Remote     PeerIpAddr      State     Type         InterfaceName
--------------------------------------------------------------------------------
8192   8192        10.10.1.1        Up         E_Dynamic         -
--------------------------------------------------------------------------------
      Total Up/DOWN Session Number : 1/0
```

4.3 LDP 与路由联动配置与管理

LDP 与路由联动是用进来流量保护，主要目的是当 MPLS 网络端到端路径故障时，确保流量切换到备份路径，尽可能地避免流量的丢失。包括 LDP 与静态路由联动，以及 LDP 与 IGP 联动，本节将具体介绍。

4.3.1 配置 LDP 与静态路由联动

存在主、备 LSP 的 MPLS 组网中，如果 LSR 之间依靠静态路由建立 LSP，则当主用链路的 LDP 会话故障（非链路故障导致）时，或者主用链路故障后再恢复时，主备 LSP 相互切换会导致流量丢失，此时采用 LDP 与静态路由联动解决此问题。

LSP 存在主备链路的组网中，如图 4-5 所示，LSR_1 和 LSR_4 之间通过静态路由连

通，LDP 在两端基于静态路由建立 LSP，正常情况下优选 LinkA。

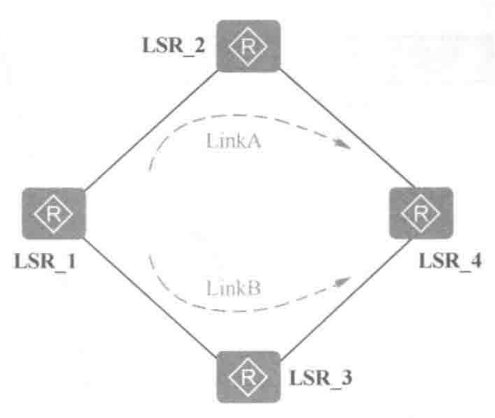


图 4-5　LDP 与静态路由联动示例

当主用链路的 LDP 会话故障（非链路故障导致），即如果 LSR_2 上的 LDP 被去使能或 LDP 出现故障，会导致 LSR_1 和 LSR_2 之间的 LDP 会话发生中断，但此时 LSR_1 和 LSR_2 之间的链路没有问题，静态路由是活跃的，路由不会切换到备份路由，而 LSP 切换到 LinkB，导致 LSR_1 和 LSR_4 之间的 MPLS 流量中断。此时如果在 LSR_1 上使能 LDP 与静态路由联动后，当 LSR_1 和 LSR_2 之间的 LDP 会话 Down 时，静态路由也自动切换到 LinkB，就可使 LSR_1 和 LSR_4 之间的 LSP 也切换到备份 LSP，保证它们之间的流量不中断。

主用链路故障后再恢复，即当 LSR_1 和 LSR_2 之间的链路失效时，LSP 会切换到 LinkB，但当 LSR_1 和 LSR_2 之间的链路恢复时，LSP 又会随静态路由切换到主用链路。此时会出现，原来的 LSP 路径无法使用，新的 LSP 还没有建立，这个时间差内，LSR_1 和 LSR_4 之间的 MPLS 流量中断。此时如果在 LSR_1 上使能 LDP 与静态路由联动，当 LSR_1 和 LSR_2 之间的 LDP 会话 Up 时，它们之间的静态路由才开始活跃，就可以保证 LSR_1 和 LSR_4 之间的流量不中断。

配置 LDP 与静态路由联动的步骤见表 4-5。

表 4-5　　配置 **LDP** 与静态路由联动的步骤

步骤	命令	说明
1	**system-view** 例如：<Huawei> **system-view**	进入系统视图
2	**ip route-static** *ip-address* { *mask* | *mask-length* } *interface-type interface-number* [*nexthop-address*] [**preference** *preference* | **tag** *tag*] * **ldp-sync** [**description** *text*] 例如：[Huawei] **ip route-static** 10.1.1.2 32 **ldp-sync**	配置 LDP 与指定的静态路由联动。命令中的参数说明如下。 • *ip-address*：指定要与 LDP 联动的静态路由的目的 IP 地址，通常为某 LSR 的 LSR ID 对应的 IP 地址。 • *mask* | *mask-length*　：指定要与 LDP 联动的静态路由的子网掩码或子网掩码长度。 • *nterface-type interface-number*：指定要与 LDP 联动的静态路由的出接口。 • *nexthop-address*：可选参数，指定要与 LDP 联动的静态路由的下一跳 IP 地址。**如果不指定本参数，则以出接口的 IP 地址作为下一跳 IP 地址**。 • **preference** *preference*：可多选可选参数，指定要与 LDP 联动的静态路由的优先级，整数形式，取值范围是 1～255。缺省值是 60。 • **tag** *tag*：可多选可选参数，指定要与 LDP 联动的静态路由的标记，整数形式，取值范围是 1～4294967295。缺省值是 0。 • **description** *text*：可选参数，LDP 与指定静态路由联动的描述。 缺省情况下，未使能 LDP 与静态路由联动功能，可用 **undo ip route-static** *ip-address* { *mask* | *mask-length* } *interface-type interface-number* [*nexthop-address*] [**preference** *preference* | **tag** *tag*] * **ldp-sync** 命令取消 LDP 与指定静态路由的联动

（续表）

步骤	命令	说明
3	**interface** *interface-type interface-number* 例如：[Huawei] **interface** gigabitethernet 1/0/0	（可选）进入静态路由主用链路的出接口视图
4	**static-route timer ldp-sync hold-down** { *timer* | **infinite** } 例如：[Huawei-GigabitEthernet 1/0/0] **static-route timer ldp-sync hold-down** 20	（可选）设置静态路由不活跃等待 LDP 会话建立的时间间隔。系统不支持在 Loopback、二层以太网接口和 NULL 接口下配置 hold-down 定时器。命令中的参数和选项说明如下。 • *timer*：二选一参数，指定静态路由不活跃等待 LDP 会话建立的时间间隔，整数形式，取值范围是 0～65535，单位是 s。当 hold-down 定时器为 0s 时，关闭该接口下的 LDP 与静态路由联动功能。 • **infinite**：二选一选项，指定定时器永远不超时。只有在 LDP 会话建立后，静态路由才活跃，MPLS 流量才进行切换。 缺省情况下，hold-down 定时器的值是 10s，可用 **undo static-route timer ldp-sync hold-down** 命令恢复为缺省配置

4.3.2 LDP 和静态路由联动配置示例

如图 4-6 所示，LSRA 有分别经过 LSRB 和 LSRC 到 LSRD 的静态路由，并基于静态路由建立了 LDP 会话，其中 LinkA 为主用链路，LinkB 为备用链路。要求 LinkA 上的 LDP 会话中断或者 LinkA 发生故障再恢复的情况下，保证 MPLS 流量不中断。

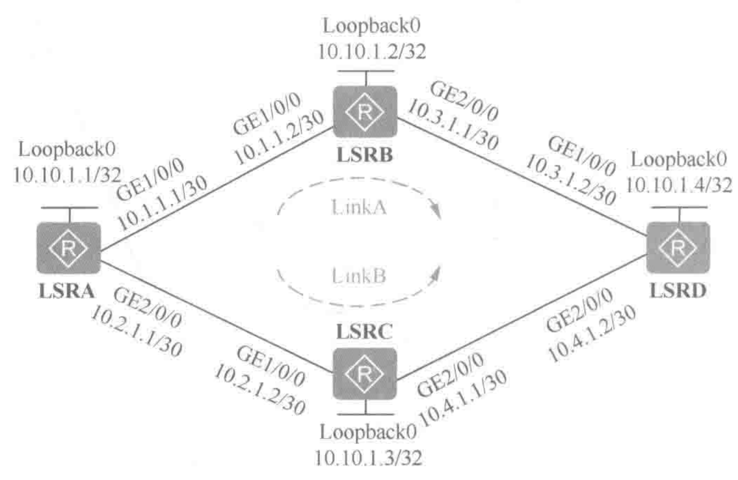


图 4-6 LDP 和静态路由联动配置示例的拓扑结构

1. 基本配置思路分析

因为本示例各 LSR 之间是通过静态路由实现互通的（其实这样配置是很少的），所以可以通过配置 LDP 和静态路由联动，调整 hold-down 定时器，实现当主用链路上的 LDP 会话中断，或者当主用链路发生故障再恢复时能保证 MPLS 流量不中断需求。

因为 LDP 与静态路由联动是扩展功能，所以首先需要配置 LDP 的必选基本功能，故本示例的基本配置思路如下。

（1）在各 LSR 上配置各接口（包括 Loopback 接口）的 IP 地址。

（2）在各 LSR 上配置静态路由（经过 LSRB 的为主用链路，经过 LSRC 的为备用链路），实现骨干网的 IP 连通性。

（3）在各 LSR 上配置 LDP，使各 LSR 间可建立到达对方 LSR ID 所代表的主机路由的 LDP LSP。

（4）在 LSRA 和 LSRD 上分别配置 LDP 和静态路由联动功能，并设置 hold-down 定时器值为 20s，即可使故障恢复（如主用链路故障恢复）的静态路由暂时不活跃，而是在 hold-down 定时器的设定值内等待 LDP 会话建立，从而达到 LDP 与静态路由的联动。

2. 具体配置步骤

（1）配置各 LSR 各接口（包括 Loopback 接口）的 IP 地址。

LSRA 上的配置。

```
<Huawei> system-view
[Huawei] sysname LSRA
[LSRA] interface loopback 0
[LSRA-LoopBack0] ip address 10.10.1.1 32
[LSRA-LoopBack0] quit
[LSRA] interface gigabitethernet 1/0/0
[LSRA-GigabitEthernet1/0/0] ip address 10.1.1.1 30
[LSRA-GigabitEthernet1/0/0] quit
[LSRA] interface gigabitethernet 2/0/0
[LSRA-GigabitEthernet2/0/0] ip address 10.2.1.1 30
[LSRA-GigabitEthernet2/0/0] quit
```

LSRB 上的配置。

```
<Huawei> system-view
[Huawei] sysname LSRB
[LSRB] interface loopback 0
[LSRB-LoopBack0] ip address 10.10.1.2 32
[LSRB-LoopBack0] quit
[LSRB] interface gigabitethernet 1/0/0
[LSRB-GigabitEthernet1/0/0] ip address 10.1.1.2 30
[LSRB-GigabitEthernet1/0/0] quit
[LSRB] interface gigabitethernet 2/0/0
[LSRB-GigabitEthernet2/0/0] ip address 10.3.1.1 30
[LSRB-GigabitEthernet2/0/0] quit
```

LSRC 上的配置。

```
<Huawei> system-view
[Huawei] sysname LSRC
[LSRC] interface loopback 0
[LSRC-LoopBack0] ip address 10.10.1.3 32
[LSRC-LoopBack0] quit
[LSRC] interface gigabitethernet 1/0/0
[LSRC-GigabitEthernet1/0/0] ip address 10.2.1.2 30
[LSRC-GigabitEthernet1/0/0] quit
[LSRC] interface gigabitethernet 2/0/0
[LSRC-GigabitEthernet2/0/0] ip address 10.4.1.1 30
[LSRC-GigabitEthernet2/0/0] quit
```

LSRD 上的配置。

```
<Huawei> system-view
```

```
[Huawei] sysname LSRD
[LSRD] interface loopback 0
[LSRD-LoopBack0] ip address 10.10.1.4 32
[LSRD-LoopBack0] quit
[LSRD] interface gigabitethernet 1/0/0
[LSRD-GigabitEthernet1/0/0] ip address 10.3.1.2 30
[LSRD-GigabitEthernet1/0/0] quit
[LSRD] interface gigabitethernet 2/0/0
[LSRD-GigabitEthernet2/0/0] ip address 10.4.1.2 30
[LSRD-GigabitEthernet2/0/0] quit
```

（2）在各节点上配置静态路由，使网络互通。

LSRA 上配置到 LSRD 的两条优先级不同的静态路由，同时 LSRD 上也相应配置到 LSRA 的两条优先级不同的静态路由。

LSRA 上的配置。

```
[LSRA] ip route-static 10.10.1.2 32 10.1.1.2
[LSRA] ip route-static 10.10.1.3 32 10.2.1.2
[LSRA] ip route-static 10.3.1.1 30 10.1.1.2
[LSRA] ip route-static 10.4.1.1 30 10.2.1.2
[LSRA] ip route-static 10.10.1.4 32 10.1.1.2 preference 40   #---到达 LSRD 的主用路由
[LSRA] ip route-static 10.10.1.4 32 10.2.1.2   #---到达 LSRD 的备用路由，采用缺省优先级值 60
```

配置 LSRB。

```
[LSRB] ip route-static 10.10.1.1 32 10.1.1.1
[LSRB] ip route-static 10.10.1.4 32 10.3.1.2
```

配置 LSRC。

```
[LSRC] ip route-static 10.10.1.1 32 10.2.1.1
[LSRC] ip route-static 10.10.1.4 32 10.4.1.2
```

配置 LSRD。

```
[LSRD] ip route-static 10.10.1.2 32 10.3.1.1
[LSRD] ip route-static 10.10.1.3 32 10.4.1.1
[LSRD] ip route-static 10.1.1.2 30 10.3.1.1
[LSRD] ip route-static 10.2.1.2 30 10.4.1.1
[LSRD] ip route-static 10.10.1.1 32 10.3.1.1 preference 40   #---到达 LSRA 的主用路由
[LSRD] ip route-static 10.10.1.1 32 10.4.1.1   #---到达 LSRA 的备用路由，采用缺省优先级值 60
```

以上配置完成后，在各节点上执行 **display ip routing-table protocol static** 命令可以查看到所配置的静态路由。以下是在 LSRA 上执行该命令的输出示例。

```
[LSRA] display ip routing-table protocol static
Route Flags: R - relay, D - download to fib
------------------------------------------------------------------------------
Public routing table : Static
         Destinations : 5          Routes : 6          Configured Routes : 6

Static routing table status : <Active>
         Destinations : 5          Routes : 5

Destination/Mask    Proto   Pre  Cost      Flags NextHop         Interface

      10.10.1.2/32  Static  60   0           RD   10.1.1.2        GigabitEthernet1/0/0
      10.10.1.3/32  Static  60   0           RD   10.2.1.2        GigabitEthernet2/0/0
      10.10.1.4/32  Static  40   0           RD   10.1.1.2        GigabitEthernet1/0/0
       10.3.1.0/30  Static  60   0           RD   10.1.1.2        GigabitEthernet1/0/0
       10.4.1.0/30  Static  60   0           RD   10.2.1.2        GigabitEthernet2/0/0
```

```
Static routing table status : <Inactive>
         Destinations : 1          Routes : 1

Destination/Mask    Proto   Pre  Cost      Flags NextHop         Interface

      10.10.1.4/32  Static  60   0           D   10.2.1.2        GigabitEthernet2/0/0
```

从中可以看到，除了有 5 条活跃（Active）状态的静态路由外，还有一条非活跃（Inactive）状态的静态路由，那就是到达 LSRD 的备用静态路由。

（3）在各 LSR 上使能 MPLS LDP 能力，建立 LDP LSP。

LSRA 上的配置。

```
[LSRA] mpls lsr-id 10.10.1.1
[LSRA] mpls
[LSRA-mpls] quit
[LSRA] mpls ldp
[LSRA-mpls-ldp] quit
[LSRA] interface gigabitethernet 1/0/0
[LSRA-GigabitEthernet1/0/0] mpls
[LSRA-GigabitEthernet1/0/0] mpls ldp
[LSRA-GigabitEthernet1/0/0] quit
[LSRA] interface gigabitethernet 2/0/0
[LSRA-GigabitEthernet2/0/0] mpls
[LSRA-GigabitEthernet2/0/0] mpls ldp
[LSRA-GigabitEthernet2/0/0] quit
```

LSRB 上的配置。

```
[LSRB] mpls lsr-id 10.10.1.2
[LSRB] mpls
[LSRB-mpls] quit
[LSRB] mpls ldp
[LSRB-mpls-ldp] quit
[LSRB] interface gigabitethernet 1/0/0
[LSRB-GigabitEthernet1/0/0] mpls
[LSRB-GigabitEthernet1/0/0] mpls ldp
[LSRB-GigabitEthernet1/0/0] quit
[LSRB] interface gigabitethernet 2/0/0
[LSRB-GigabitEthernet2/0/0] mpls
[LSRB-GigabitEthernet2/0/0] mpls ldp
[LSRB-GigabitEthernet2/0/0] quit
```

LSRC 上的配置。

```
[LSRC] mpls lsr-id 10.10.1.3
[LSRC] mpls
[LSRC-mpls] quit
[LSRC] mpls ldp
[LSRC-mpls-ldp] quit
[LSRC] interface gigabitethernet 1/0/0
[LSRC-GigabitEthernet1/0/0] mpls
[LSRC-GigabitEthernet1/0/0] mpls ldp
[LSRC-GigabitEthernet1/0/0] quit
[LSRC] interface gigabitethernet 2/0/0
[LSRC-GigabitEthernet2/0/0] mpls
[LSRC-GigabitEthernet2/0/0] mpls ldp
[LSRC-GigabitEthernet2/0/0] quit
```

LSRD 上的配置。

```
[LSRD] mpls lsr-id 10.10.1.4
[LSRD] mpls
[LSRD-mpls] quit
[LSRD] mpls ldp
[LSRD-mpls-ldp] quit
[LSRD] interface gigabitethernet 1/0/0
[LSRD-GigabitEthernet1/0/0] mpls
[LSRD-GigabitEthernet1/0/0] mpls ldp
[LSRD-GigabitEthernet1/0/0] quit
[LSRD] interface gigabitethernet 2/0/0
[LSRD-GigabitEthernet2/0/0] mpls
[LSRD-GigabitEthernet2/0/0] mpls ldp
[LSRD-GigabitEthernet2/0/0] quit
```

以上配置完成后，在各节点上执行 **display mpls ldp session** 命令可以看到它们的 LDP Session 已经建立（状态为 **Operational**）。以下是在 LSRA 上执行该命令的输出示例，从中可以看到 LSRA 分别与 LSRB 和 LSRC 成功建立了本 LDP 会话。

```
[LSRA] display mpls ldp session

 LDP Session(s) in Public Network
 Codes: LAM(Label Advertisement Mode), SsnAge Unit(DDDD:HH:MM)
 A '*' before a session means the session is being deleted.
 ------------------------------------------------------------------------------
 PeerID              Status       LAM   SsnRole   SsnAge        KASent/Rcv
 ------------------------------------------------------------------------------
 10.10.1.2:0         Operational DU     Passive   0000:00:00    1/1
 10.10.1.3:0         Operational DU     Passive   0000:00:02    12/12
 ------------------------------------------------------------------------------
 TOTAL: 2 session(s) Found.
```

（4）在 LSRA 和 LSRD 上分别配置 LDP 和静态路由联动功能，调整 hold-down 定时器值为 20s。

LSRA 上的配置。

```
[LSRA] ip route-static 10.10.1.4 32 gigabitethernet 1/0/0 10.1.1.2 ldp-sync
[LSRA] interface gigabitethernet 1/0/0
[LSRA-GigabitEthernet1/0/0] static-route timer ldp-sync hold-down 20
[LSRA-GigabitEthernet1/0/0] quit
```

配置 LSRD。

```
[LSRD] ip route-static 10.10.1.1 32 gigabitethernet 1/0/0 10.3.1.1 ldp-sync
[LSRD] interface gigabitethernet 1/0/0
[LSRD-GigabitEthernet1/0/0] static-route timer ldp-sync hold-down 20
[LSRD-GigabitEthernet1/0/0] quit
```

3. 配置结果验证

以上配置完成后，在 LSRA 上查看使能了 LDP 和静态路由联动功能的静态路由出接口的状态信息。

```
[LSRA] display static-route ldp-sync
Total number of routes enable Ldp-Sync: 1
 ------------------------------------------------------------------------------
Interface GigabitEthernet1/0/0
Enable ldp-sync static routes number: 1
Static-route ldp-sync holddown timer: 20s
```

```
Sync state: Normal
Dest = 10.10.1.4, Mask = 32, NextHop = 10.1.1.2.
------------------------------------------------------------------------
```

可以看到，LDP 和静态路由联动功能已经配置（状态为 **Normal**）。这样就可实现当主用链路（LinkA）的 LDP 会话中断时，为了保证静态路由与 LSP 的联动，流量立即切换到备用链路（LinkB）来保证流量不中断。当主用链路发生故障再恢复时，下一跳为 10.1.1.2 的静态路由并不会马上被优选。只有等到 hold-down 定时器超时（20s），主用链路的 LDP 会话建立后，才会看到下一跳为 10.1.1.2 的静态路由活跃，达到静态路由和 LDP 的联动，从而保证 MPLS 流量不中断。

4.3.3 配置 LDP 与 IGP 联动

LDP 与 IGP 联动主要用在存在主备 LSP 的 MPLS 组网中，LSR 之间依靠 IGP 路由协议（主要是 OSPF 和 IS-IS 协议）建立 LSP。但由于 LDP 的收敛速度依赖于 IGP 路由的收敛，即 LDP 的收敛速度比 IGP 路由的收敛速度慢，因此在主、备链路的组网中使用 MPLS LDP 会存在以下问题。

■ 当主用链路发生故障时，IGP 路由和 LSP 均切换到备用链路上（常通过 LDP FRR 实现，本书不作介绍，可参见相关文档）。但当主用链路从故障中恢复时，IGP 会先于 LDP 切换回主用链路，因此会造成 LSP 流量丢失。

■ 当主用链路 IGP 运行正常，但主用链路节点间的 LDP 会话发生故障时，主用链路的 LSP 会被删除，而 IGP 路由会仍然使用主用链路。同时，由于备用链路不存在 IGP 优选路由，故 LSP 无法在备用链路建立，导致 LSP 流量丢失。

此时可通过使能 LDP 与 IGP 联动功能，在主用链路的 LDP 会话故障（非链路故障导致）时，或者主用链路故障后再恢复时，用来解决主备 LSP 相互切换导致的流量丢失问题。在配置时，需在主、备链路的分叉节点和主用链路上的 LDP 邻居节点之间的链路两端接口上同时配置。

LDP 与 IGP 联动所包括的配置任务如下，但在配置 LDP 与 IGP 联动之前要先完成 LDP 本地会话配置。

■ 使能 LDP 与 IGP 联动功能。

■（可选）阻止接口上运行 LDP 与 IS-IS 联动功能。

■（可选）设置 hold-down 定时器的值。

■（可选）设置 hold-max-cost 定时器的值。

■（可选）设置 Delay 定时器的值。

1. 使能 LDP 与 IGP 联动功能

使能 LDP 与 IGP 联动有两种方式。

■ 在接口视图中使能 LDP 与 IGP 联动；

对应接口下 LDP 和 IGP 联动功能将使能，适合只有少量接口需使能 LDP 和 IGP 联动的场景。

■ 在 IGP 进程中使能 LDP 与 IGP 联动；

对应 IGP 进程下的接口都将自动使能 LDP 和 IGP 联动功能。如果同一节点上有很多接口都需要使能 LDP 与 IGP 联动，则推荐此种方式进行配置。目前仅 LDP 和 IS-IS

联动功能支持这种使能方式。

（1）LDP 与 OSPF 联动

使能 LDP 与 OSPF 联动功能的方法是在主、备链路的分叉节点和主用链路上的 LDP 邻居节点之间的链路两端接口上执行 **ospf ldp-sync** 命令。缺省情况下，接口上未使能 LDP 和 OSPF 联动功能，可用 **undo ospf ldp-sync** 命令去使能接口 LDP 和 OSPF 联动功能。

（2）LDP 与 IS-IS 联动

IGP 与 IS-IS 联动功能的使能，可在 IS-IS 接口或 IS-IS 进程下分别进行，具体的配置步骤如表 4-6 所示。接口下的配置优先级高于 IS-IS 进程下的配置，同时配置且当两者的配置不一致时，接口下的配置生效。

表 4-6　配置 LDP 与 IS-IS 路由联动的步骤

步骤	命令	说明
1	**system-view** 例如：<Huawei> **system-view**	进入系统视图
方法一：在 IS-IS 接口下使能 LDP 与 IS-IS 联动		
2	**interface** *interface-type interface-number* 例如：[Huawei] **interface** gigabitethernet 1/0/0	进入接口视图
	isis enable [*process-id*] 例如：[Huawei-GigabitEthernet1/0/0] **isis enable** 1	使能 IS-IS 协议。可选参数 *process-id* 代表要启动的 IS-IS 路由进程号，整数类型，取值范围是 1～65535，缺省值是 1。**一个接口只能与一个 IS-IS 进程相关联。** 缺省情况下，接口上未使能 IS-IS 功能，可用 **undo isis enable** 命令用来在接口上去使能 IS-IS 功能并取消与 IS-IS 进程号的关联
	isis ldp-sync 例如：[Huawei-GigabitEthernet1/0/0] **ospf ldp-sync**	使能接口 LDP 与 IS-IS 联动功能。需要在主、备链路的分叉节点和主用链路上的 LDP 邻居节点之间的链路两端接口上同时配置。 缺省情况下，接口上未使能 LDP 与 IS-IS 联动功能，可用 **undo ospf ldp-sync** 命令去使能接口的 LDP 和 OSPF 联动功能
方法二：在 IS-IS 路由进程下配置 LDP 与 IS-IS 联动		
2	**isis** [*process-id*] 例如：[Huawei] **isis** 2	使能 IS-IS 协议，进入 IS-IS 进程视图。可选参数 *process-id* 用来指定一个 IS-IS 进程，整数形式，取值范围是 1～65535。如果不指定该参数，则系统默认的进程为 1。 缺省情况下，未使能 IS-IS 协议，可用 **undo isis** *process-id* 命令用来去使能 IS-IS 协议
3	**ldp-sync enable** [**mpls-binding-only**] 例如：[Huawei-isis-2] **ldp-sync enable**	使能以上 IS-IS 进程下所有接口的 LDP 和 IS-IS 联动功能。可选项 **mpls-binding-only** 用来指定只有使能 MPLS LDP 接口才使能 LDP 和 IS-IS 联动功能。 缺省情况下，IS-IS 进程下的接口没有使能 LDP 和 IS-IS 联动功能，可用 **undo ldp-sync enable** 命令去使能 IS-IS 进程下所有接口的 LDP 和 IS-IS 联动功能

2. 阻止接口上运行 LDP 与 IS-IS 联动功能

对 IS-IS 进程执行 **ldp-sync enable** 命令后，该 IS-IS 进程下的所有接口都将使能 LDP

和 IS-IS 联动功能。但是，对于连接着重要业务节点的 IS-IS 接口，运行 LDP 和 IS-IS 联动功能可能造成如下问题：当链路正常而 LDP 会话出现故障时，IS-IS 将在当前节点的 LSP（Link State PDU，链路状态 PDU）中通告最大开销值，导致 IS-IS 路由不再优选当前链路（成了备用链路上的节点设备），从而影响重要业务的运行。

为了避免出现上述问题，可以阻止指定的 IS-IS 接口运行 LDP 和 IS-IS 联动功能。配置的方法是在对应的 IS-IS 接口视图下执行 **isis ldp-sync block** 命令，阻止该接口上运行 LDP 与 IS-IS 联动功能。缺省情况下，接口上不阻止 LDP 与 IS-IS 联动功能，可用 **undo isis ldp-sync block** 命令恢复为缺省配置。

3. 设置 hold-down 定时器的值

在使能 LDP 和 IGP 联动功能后，当主用链路物理故障恢复时，IGP 进入 hold-down 状态并启动 hold-down 定时器。在 hold-down 定时器超时之前，IGP 都不会建立邻居关系，以便等待 LDP 会话建立，达到 LDP 和 IGP 同步回切到主用链路上的目的。

说明 如果配置的是 LDP 与 IS-IS 联动，则既可以在指定 IS-IS 接口下设置 hold-down 定时器的值，也可以在 IS-IS 视图下统一设置所有 IS-IS 接口的 hold-down 定时器的值。接口下的配置优先级高于 IS-IS 进程下的配置，当两者不一致时，接口下的配置生效。

LDP 和 IGP 联动 Hold-down 定时器的配置方法见表 4-7。

表 4-7　配置 LDP 与 IGP 路由联动 hold-down 定时器的步骤

步骤	命令	说明
1	**system-view** 例如：<Huawei> **system-view**	进入系统视图
在 LDP 与 OSPF 联动中的 hold-down 定时器配置		
2	**interface** *interface-type interface-number* 例如：[Huawei] **interface** gigabitethernet 1/0/0	进入接口视图
3	**ospf timer ldp-sync hold-down** *value* 例如：[Huawei-GigabitEthernet1/0/0] **ospf timer ldp-sync hold-down** 15	设置接口不建立 OSPF 邻居而等待 LDP 会话建立的时间间隔，数形式，取值范围是 0～65535，单位是 s。 缺省情况下，不建立 OSPF 邻居而等待 LDP 会话建立的时间间隔是 10s，可用 **undo ospf timer ldp-sync hold-down** 命令恢复缺省配置
在 LDP 与 IS-IS 联动中指定 IS-IS 接口的 hold-down 定时器配置		
2	**interface** *interface-type interface-number* 例如：[Huawei] **interface** gigabitethernet 1/0/0	进入接口视图
3	**isis timer ldp-sync hold-down** *value* 例如：[Huawei-GigabitEthernet1/0/0] **isis timer ldp-sync hold-down** 300	设置接口不建立 IS-IS 邻居而等待 LDP 会话建立的时间间隔，整数形式，取值范围是 0～65535，单位是 s。 缺省情况下，hold-down 定时器的值是 10s，可用 **undo isis timer ldp-sync hold-down** 命令恢复为缺省配置

（续表）

步骤	命令	说明
在 LDP 与 IS-IS 联动中所有 IS-IS 接口的 hold-down 定时器配置		
2	**isis** [*process-id*] 例如：[Huawei] **isis** 100	进入 IS-IS 进程视图
3	**timer ldp-sync hold-down** *value* 例如：[Huawei-isis-100] **timer ldp-sync hold-down** 15	设置所有使能 IS-IS 的接口为了等待 LDP 会话建立而保持 hold-down 状态的时间，整数形式，取值范围是 0～65535，单位是 s。 缺省情况下，接口为了等待 LDP 会话建立而保持 hold-down 状态的时间是 10s，可用 **undo timer ldp-sync hold-down** 命令恢复为缺省配置

4. 设置 hold-max-cost 定时器的值

在使能 LDP 和 IGP 联动功能后，如果主用链路 LDP 会话发生故障，但 IGP 协议正常，为了使 IGP 和 LDP 同步切换到备用链路，IGP 协议会在本节点的 LSP（Link State PDU）中通告最大开销值。通过设置 hold-max-cost 定时器的值，可以调整 IGP 通告最大开销值的持续时间。

设置 hold-max-cost 定时器的值有两种方式，具体配置方法见表 4-8。

（1）在接口视图中设置 hold-max-cost 定时器的值

对应接口下设置 hold-max-cost 定时器的值，适合只有少量接口需设置 hold-max-cost 定时器的值的场景。

（2）在 IGP 进程中设置 hold-max-cost 定时器的值

对应 IGP 进程下的接口都将自动设置成该值。如果同一节点上有很多接口都需要设置 hold-max-cost 定时器的值，则推荐此种方式进行配置。

说明 如果配置的是 LDP 与 IS-IS 联动，则既可以在指定 IS-IS 接口下设置 hold-max-cost 定时器的值，也可以在 IS-IS 视图下统一设置所有 IS-IS 接口的 hold-max-cost 定时器的值。接口下的配置优先级高于 IS-IS 进程下的配置，当两者不一致时，接口下的配置生效。

表 4-8 配置 LDP 与 IGP 路由联动 hold-max-cost 定时器的步骤

步骤	命令	说明
1	**system-view** 例如：<Huawei> **system-view**	进入系统视图
在 LDP 与 OSPF 联动中的 hold-max-cost 定时器配置		
2	**interface** *interface-type interface-number* 例如：[Huawei] **interface** gigabitethernet 1/0/0	进入接口视图
3	**ospf timer ldp-sync hold- max-cost** { *value* | **infinite** } 例如：[Huawei-GigabitEthernet 1/0/0] **ospf timer ldp-sync hold-max-cost** 10	配置 OSPF 在本地节点的 LSA（Link State Advertisement）中保持通告最大开销值的时间。命令中的参数和选项说明如下。 • *value*：二选参数，指定 OSPF 在本地设备的 LSA 中保持通告最大开销值的时间，整数形式，取值范围是 0～65535，单位是秒

（续表）

步骤	命令	说明
3	**ospf timer ldp-sync hold-max-cost** { *value* | **infinite** } 例如：[Huawei-GigabitEthernet 1/0/0] **ospf timer ldp-sync hold-max-cost** 10	• **infinite**：二选一选项，指定在LDP会话重新建立之前，OSPF在本地设备的LSA中永久通告最大开销值 缺省情况下，hold-max-cost定时器的值是10秒，可用**undo ospf timer ldp-sync hold-max-cost**命令恢复为缺省配置
在LDP与IS-IS联动中指定IS-IS接口的hold-max-cost定时器配置		
2	**interface** *interface-type interface-number* 例如：[Huawei] **interface** gigabitethernet 1/0/0	进入接口视图
3	**isis timer ldp-sync hold-max-cost** { *value* | **infinite** } 例如：[Huawei-GigabitEthernet 1/0/0] **isis timer ldp-sync hold-max-cost** 60	配置IS-IS在本地设备的LSP（Link State PDU）中保持通告最大开销值的时间。命令中的参数和选项说明如下。 • *value*：二选一参数，指定IS-IS在本地设备的IS-IS LSP（Link State PDU）中保持通告最大开销值的时间，整数形式，取值范围是0～65535，单位是秒 • **infinite**：二选一选项，指定在LDP会话重新建立之前，IS-IS在本地设备的LSP中永久通告最大开销值 缺省情况下，IS-IS在本地设备的LSP中保持通告最大开销值的时间是10秒，可用**undo isis timer ldp-sync hold-max-cost**命令恢复为缺省配置
在LDP与IS-IS联动中所有IS-IS接口的hold-max-cost定时器配置		
2	**isis** [*process-id*] 例如：[Huawei] **isis** 100	进入IS-IS进程视图
3	**timer ldp-sync hold-max-cost** { **infinite** | *interval* } 例如：[Huawei-isis-100] **timer ldp-sync hold-max-cost** 60	配置所有使能LDP和IS-IS同步功能的接口保持通告最大开销值的时间，参见和选项说明参见本表前面在IS-IS接口下的配置说明 缺省情况下，所有使能LDP和IS-IS同步功能的接口保持通告最大开销值的时间是10s，可用**undo timer ldp-sync hold-max-cost**命令恢复为缺省情况

说明　根据不同组网需要，可选择参数进行配置。

■ 如果组网中IGP仅承载LDP业务，要使IGP的选路和LDP LSP始终保持一致，需选择**infinite**选项。

■ 如果组网中IGP承载了包括LDP在内的多种业务时，要使LDP会话的中断不影响IGP的正常选路和其他业务，可配置*value*参数。

5. 设置Delay定时器的值

故障链路的LDP会话重新建立以后，LDP会启动Delay定时器等待LSP的建立，当Delay定时器超时以后，LDP会通知IGP联动流程结束。一般可直接采用缺省值。

这个Delay定时器的配置方法是在具体接口下通过**mpls ldp timer igp-sync-delay** *value*命令进行的。缺省情况下，LDP会话建立后等待LSP建立的时间间隔是10s，可用**undo mpls ldp timer igp-sync-delay**命令恢复为缺省配置。

4.3.4 LDP 与 IGP 联动管理命令

配置好上面各小节的 LDP 与 IGP 路由联动的相关功能和参数后，可在任意视图下通过以下 **display** 命令查看相关配置信息。

■ **display ospf ldp-sync interface** { **all** | *interface-type interface-number* }：查看配置了 LDP 与 OSPF 联动功能的接口的同步信息。

■ **display isis** [*process-id*] **ldp-sync interface**：查看配置了 LDP 与 IS-IS 联动功能的接口的同步信息。

■ **display rm interface** [*interface-type interface-number* | **vpn-instance** *vpn-instance-name*]：查看接口的路由管理信息。

4.3.5 LDP 与 OSPF 联动配置示例

如图 4-7 所示，MPLS 骨干网络包括 P1、P2、P3、PE2 四个节点，各设备间运行 OSPF 路由协议。PE1 到 PE2 之间建立两条 LSP 链路，PE1→P1→P2→PE2 为主用链路，PE1→P1→P3→PE2 为备用链路。当主用链路故障恢复时，由于 OSPF 路由比 LDP 收敛速度快，OSPF 会先于 LDP 切换回主用链路，因此造成 LSP 流量丢失。现要求能够解决在 LSP 存在主、备链路的组网中 LSP 流量丢失问题。

1. 基本配置思路分析

本示例存在主、备 MPLS 链路，且 MPLS 骨干网采用 OSPF 路由协议，为了防止 LSP 流量的丢失，可配置 LDP 与 OSPF 联动。在主、备链路的分叉节点 P1 和主用链路上的 LDP 邻居节点 P2 之间的链路两端接口上使能 LDP 与 OSPF 联动功能。还可根据需要，在主、备链路的分叉节点 P1 和主用链路上的 LDP 邻居节点 P2 之间的链路两端接口上设置定时器 hold-down、hold-max-cost 和 delay 的值。

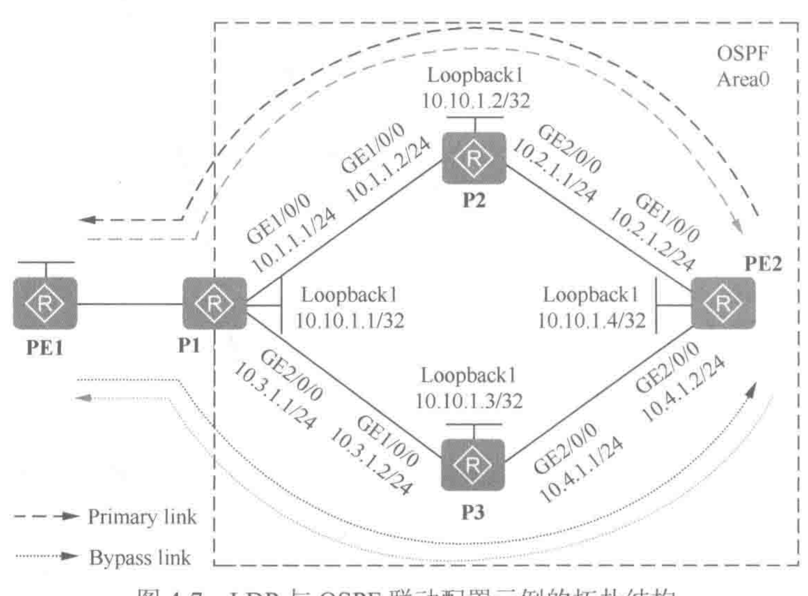


图 4-7 LDP 与 OSPF 联动配置示例的拓扑结构

在进行以上配置之前，先要完成 LDP 必选项基本功能配置，并配置好 MPLS 骨干

网的 OSPF 路由。下面是本示例的基本配置思路。

（1）在各 LSR 上配置各接口（包括 Loopback 接口）的 IP 地址。

（2）在各 LSR 上配 OSPF 态路由（通过增加 P1 的 GE2/0/0 接口的开销值，把经过 P3 的链路配置为备用链路），实现骨干网的 IP 连通性。

（3）在各 LSR 上配置 LDP，使各 LSR 间可建立到达对方 LSR ID 所代表的主机路由的 LDP LSP。

（4）在主用链路 P1 和 P2 上配置 LDP 与 OSPF 联动，并且可根据需要调整几个定时器的取值。

2. 具体配置步骤

（1）配置 MPLS 骨干网各节点上各接口的 IP 地址。

P1 上的配置。

```
<Huawei> system-view
[Huawei] sysname P1
[P1] interface loopback 1
[P1-LoopBack1] ip address 10.10.1.1 32
[P1-LoopBack1] quit
[P1] interface gigabitethernet 1/0/0
[P1-GigabitEthernet1/0/0] ip address 10.1.1.1 24
[P1-GigabitEthernet1/0/0] quit
[P1] interface gigabitethernet 2/0/0
[P1-GigabitEthernet2/0/0] ip address 10.3.1.1 24
[P1-GigabitEthernet2/0/0] quit
```

P2 上的配置。

```
<Huawei> system-view
[Huawei] sysname P2
[P2] interface loopback 1
[P2-LoopBack1] ip address 10.10.1.2 32
[P2-LoopBack1] quit
[P2] interface gigabitethernet 1/0/0
[P2-GigabitEthernet1/0/0] ip address 10.1.1.2 24
[P2-GigabitEthernet1/0/0] quit
[P2] interface gigabitethernet 2/0/0
[P2-GigabitEthernet2/0/0] ip address 10.2.1.1 24
[P2-GigabitEthernet2/0/0] quit
```

P3 上的配置。

```
<Huawei> system-view
[Huawei] sysname P3
[P3] interface loopback 1
[P3-LoopBack1] ip address 10.10.1.3 32
[P3-LoopBack1] quit
[P3] interface gigabitethernet 1/0/0
[P3-GigabitEthernet1/0/0] ip address 10.3.1.2 24
[P3-GigabitEthernet1/0/0] quit
[P3] interface gigabitethernet 2/0/0
[P3-GigabitEthernet2/0/0] ip address 10.4.1.1 24
[P3-GigabitEthernet2/0/0] quit
```

PE2 上的配置。

```
<Huawei> system-view
[Huawei] sysname PE2
```

```
[PE2] interface loopback 1
[PE2-LoopBack1] ip address 10.10.1.4 32
[PE2-LoopBack1] quit
[PE2] interface gigabitethernet 1/0/0
[PE2-GigabitEthernet1/0/0] ip address 10.2.1.2 24
[PE2-GigabitEthernet1/0/0] quit
[PE2] interface gigabitethernet 2/0/0
[PE2-GigabitEthernet2/0/0] ip address 10.4.1.2 24
[P2-GigabitEthernet2/0/0] quit
```

（2）配置 OSPF 协议发布各节点接口所连网段和 LSR ID 的主机路由。

P1 上的配置。

```
[P1] ospf 1
[P1-ospf-1] area 0
[P1-ospf-1-area-0.0.0.0] network 10.10.1.1 0.0.0.0
[P1-ospf-1-area-0.0.0.0] network 10.1.1.0 0.0.0.255
[P1-ospf-1-area-0.0.0.0] network 10.3.1.0 0.0.0.255
[P1-ospf-1-area-0.0.0.0] quit
[P1-ospf-1] quit
```

P2 上的配置。

```
[P2] ospf 1
[P2-ospf-1] area 0
[P2-ospf-1-area-0.0.0.0] network 10.10.1.2 0.0.0.0
[P2-ospf-1-area-0.0.0.0] network 10.1.1.0 0.0.0.255
[P2-ospf-1-area-0.0.0.0] network 10.2.1.0 0.0.0.255
[P2-ospf-1-area-0.0.0.0] quit
[P2-ospf-1] quit
```

P3 上的配置。

```
[P3] ospf 1
[P3-ospf-1] area 0
[P3-ospf-1-area-0.0.0.0] network 10.10.1.3 0.0.0.0
[P3-ospf-1-area-0.0.0.0] network 10.3.1.0 0.0.0.255
[P3-ospf-1-area-0.0.0.0] network 10.4.1.0 0.0.0.255
[P3-ospf-1-area-0.0.0.0] quit
[P3-ospf-1] quit
```

PE1 上的配置。

```
[PE1] ospf 1
[PE1-ospf-1] area 0
[PE1-ospf-1-area-0.0.0.0] network 10.10.1.4 0.0.0.0
[PE1-ospf-1-area-0.0.0.0] network 10.2.1.0 0.0.0.255
[PE1-ospf-1-area-0.0.0.0] network 10.4.1.0 0.0.0.255
[PE1-ospf-1-area-0.0.0.0] quit
[PE1-ospf-1] quit
```

配置 P1 的 GE2/0/0 接口的 cost 值为 1000（其他接口保持缺省的开销值，GE 缺省的开销值为 1，开销值越大，优先级越低），使得 P1→P3→PE2 链路为备用链路。

```
[P1] interface gigabitethernet 2/0/0
[P1-GigabitEthernet2/0/0] ospf cost 1000
[P1-GigabitEthernet2/0/0] quit
```

上述配置完成后，在各节点上执行 **display ip routing-table** 命令，可以看到相互之间都学到了到达彼此的路由，且 P1 到 PE1（10.10.1.4/32）路由的出接口为 GE1/0/0。以下是在 P1 上执行该命令的输出示例。

```
[P1] display ip routing-table
```

```
Route Flags: R - relay, D - download to fib
------------------------------------------------------------------------------
Routing Tables: Public
         Destinations : 16        Routes : 16
Destination/Mask    Proto   Pre  Cost      Flags NextHop         Interface
      10.10.1.1/32  Direct  0    0           D   127.0.0.1       LoopBack1
      10.10.1.2/32  OSPF    10   1           D   10.1.1.2        GigabitEthernet1/0/0
      10.10.1.3/32  OSPF    10   3           D   10.1.1.2        GigabitEthernet1/0/0
      10.10.1.4/32  OSPF    10   2           D   10.1.1.2        GigabitEthernet1/0/0
       10.1.1.0/24  Direct  0    0           D   10.1.1.1        GigabitEthernet1/0/0
       10.1.1.1/32  Direct  0    0           D   127.0.0.1       GigabitEthernet1/0/0
     10.1.1.255/32  Direct  0    0           D   127.0.0.1       GigabitEthernet1/0/0
       10.2.1.0/24  OSPF    10   2           D   10.1.1.2        GigabitEthernet1/0/0
       10.3.1.0/24  Direct  0    0           D   10.3.1.1        GigabitEthernet2/0/0
       10.3.1.1/32  Direct  0    0           D   127.0.0.1       GigabitEthernet2/0/0
     10.3.1.255/32  Direct  0    0           D   127.0.0.1       GigabitEthernet2/0/0
       10.4.1.0/24  OSPF    10   3           D   10.1.1.2        GigabitEthernet1/0/0
        127.0.0.0/8  Direct  0    0           D   127.0.0.1       InLoopBack0
       127.0.0.1/32  Direct  0    0           D   127.0.0.1       InLoopBack0
127.255.255.255/32  Direct  0    0           D   127.0.0.1       InLoopBack0
255.255.255.255/32  Direct  0    0           D   127.0.0.1       InLoopBack0
```

（3）使能各节点全局和各接口的 MPLS 和 MPLS LDP。

\#　P1 上的配置。

```
[P1] mpls lsr-id 10.10.1.1
[P1] mpls
[P1-mpls] quit
[P1] mpls ldp
[P1-mpls-ldp] quit
[P1] interface gigabitethernet 1/0/0
[P1-GigabitEthernet1/0/0] mpls
[P1-GigabitEthernet1/0/0] mpls ldp
[P1-GigabitEthernet1/0/0] quit
[P1] interface gigabitethernet 2/0/0
[P1-GigabitEthernet2/0/0] mpls
[P1-GigabitEthernet2/0/0] mpls ldp
[P1-GigabitEthernet2/0/0] quit
```

\#　P2 上的配置。

```
[P2] mpls lsr-id 10.10.1.2
[P2] mpls
[P2-mpls] quit
[P2] mpls ldp
[P2-mpls-ldp] quit
[P2] interface gigabitethernet 1/0/0
[P2-GigabitEthernet1/0/0] mpls
[P2-GigabitEthernet1/0/0] mpls ldp
[P2-GigabitEthernet1/0/0] quit
[P2] interface gigabitethernet 2/0/0
[P2-GigabitEthernet2/0/0] mpls
[P2-GigabitEthernet2/0/0] mpls ldp
[P2-GigabitEthernet2/0/0] quit
```

\#　P3 上的配置。

```
[P3] mpls lsr-id 10.10.1.3
[P3] mpls
```

```
[P3-mpls] quit
[P3] mpls ldp
[P3-mpls-ldp] quit
[P3] interface gigabitethernet 1/0/0
[P3-GigabitEthernet1/0/0] mpls
[P3-GigabitEthernet1/0/0] mpls ldp
[P3-GigabitEthernet1/0/0] quit
[P3] interface gigabitethernet 2/0/0
[P3-GigabitEthernet2/0/0] mpls
[P3-GigabitEthernet2/0/0] mpls ldp
[P3-GigabitEthernet2/0/0] quit
```

PE2 上的配置。

```
[PE2] mpls lsr-id 10.10.1.4
[PE2] mpls
[PE2-mpls] quit
[PE2] mpls ldp
[PE2-mpls-ldp] quit
[PE2] interface gigabitethernet 1/0/0
[PE2-GigabitEthernet1/0/0] mpls
[PE2-GigabitEthernet1/0/0] mpls ldp
[PE2-GigabitEthernet1/0/0] quit
[PE2] interface gigabitethernet 2/0/0
[PE2-GigabitEthernet2/0/0] mpls
[PE2-GigabitEthernet2/0/0] mpls ldp
[PE2-GigabitEthernet2/0/0] quit
```

上述配置完成后，相邻节点之间应该建立起 LDP 会话。在各节点上执行 **display mpls ldp session** 命令可以看到显示结果中 Status 项为“Operational”，表示相邻节点之间的本地 LDP 会话建立是成功的。以下是在 P1 上执行该命令输出示例，它已与 P2 和 P3 分别成功建立了本地 LDP 会话。

```
[P1] display mpls ldp session

 LDP Session(s) in Public Network
 Codes: LAM(Label Advertisement Mode), SsnAge Unit(DDDD:HH:MM)
 A '*' before a session means the session is being deleted.
 ------------------------------------------------------------------------------
 PeerID              Status        LAM   SsnRole   SsnAge        KASent/Rcv
 ------------------------------------------------------------------------------
 10.10.1.2:0         Operational DU      Passive   000:00:56     227/227
 10.10.1.3:0         Operational DU      Passive   000:00:56     227/227
 ------------------------------------------------------------------------------
 TOTAL: 2 session(s) Found.
```

（4）在主备链路的分叉节点 P1 和主用链路上的 LDP 邻居节点 P2 之间的链路两端接口上使能 LDP 与 OSPF 联动功能，并可根据需要调整它们的定时器 hold-down（设置接口不建立 OSPF 邻居而等待 LDP 会话建立的时间间隔，调整为 8s，缺省为 10s）、Hold-max-cost（配置 OSPF 在本地设备的 LSA 中保持通告最大开销值的时间，调整为 9s，缺省为 10s）和 Delay（配置 LDP 会话建立后等待 LSP 建立的时间间隔，调整为 6s，缺省也为 10s）的值。

P1 上的配置。

```
[P1] interface gigabitethernet 1/0/0
[P1-GigabitEthernet1/0/0] ospf ldp-sync
```

```
[P1-GigabitEthernet1/0/0] ospf timer ldp-sync hold-down 8
[P1-GigabitEthernet1/0/0] ospf timer ldp-sync hold-max-cost 9
[P1-GigabitEthernet1/0/0] mpls ldp timer igp-sync-delay 6
[P1-GigabitEthernet1/0/0] quit
```

P2 上的配置。

```
[P2] interface gigabitethernet 1/0/0
[P2-GigabitEthernet1/0/0] ospf ldp-sync
[P2-GigabitEthernet1/0/0] ospf timer ldp-sync hold-down 8
[P2-GigabitEthernet1/0/0] ospf timer ldp-sync hold-max-cost 9
[P2-GigabitEthernet1/0/0] mpls ldp timer igp-sync-delay 6
[P2-GigabitEthernet1/0/0] quit
```

3. 配置结果验证

上述配置完成后，在 P1 节点上执行 **display ospf ldp-sync** 命令，可以看到接口状态为“Sync-Achieved”，表示 LDP 和 OSPF 路由已同步，证明以上配置是正确的。

```
[P1] display ospf ldp-sync interface gigabitethernet 1/0/0
Interface GigabitEthernet1/0/0
HoldDown Timer: 8          HoldMaxCost Timer: 9
LDP State: Up              OSPF Sync State: Sync-Achieved
```

4.4 LDP FRR 配置与管理

LDP FRR（Fast Reroute，快速重路由）是由 IP 网络中的 IP FRR 技术扩展而来，专为 MPLS 网络提供快速重路由功能，实现了链路备份。FRR 就是快速重新选择新路由路径的意思，实现 LDP FRR 目的就是发现主 LSP 出现故障时，能快速地将流量切换到备份 LSP 上（前提是该备份路径的路由是通的），从而最大程度上避免流量的丢失。很显然，这需要事先有建立备份 LSP 的路径，这是通过 LDP 的自由标签保持方式来实现的，可通过获取的 Liberal 标签来建立备份 LSP。

4.4.1 LDP FRR 的两种实现方式

IP 网络中的 IP FRR 是指当物理层或链路层检测到故障时将故障消息上报至上层路由系统，并立即启用一条备份链路转发报文，可快速实现路由备份。但 IP FRR 是针对 IP 网络路由而设计的，仅可检测链路的物理层或链路层故障。在 MPLS 骨干网中，当主链路出现故障时，虽然有 IP FRR 使 IGP 路由快速收敛，切换到备份路径，但此时还得在切换后的备份路径上重新建立 LSP 才能通过 MPLS 隧道转发数据，这个过程无法避免流量丢失。另外，当 LSP 故障（非主链路故障引起）时，只能等待重新建立 LSP 后恢复流量转发，这会引起 MPLS 流量长时间中断。因此需要一种能够在 MPLS 网络中提供快速重路由的解决方案，即 LDP FRR。

当主 LDP LSP 出现故障时，LDP FRR 可通过 LDP 信令协议的自由标签保持方式（Liberal），先获取 Liberal Label，然后为该标签申请转发表项资源，并将转发信息下发到转发平面作为主 LSP 的备用转发表项。当接口故障（接口自己感知或者结合 BFD 检测）或者主 LSP 不通（结合 BFD 检测）时，可以快速地将流量切换至备份路径，从而实现了对主 LSP 的保护。

LDP FRR 对 LSP 的保护有两种方式。

（1）Manual LDP FRR

手动配置的 LDP FRR 需要使用命令来指定建立的备份 LSP 的出接口和下一跳。此时，当 LDP 获取的 Liberal Label 中来源匹配指定的出接口和下一跳的时候，就能够建立备份 LSP 并下发转发表项。

（2）Auto LDP FRR

Auto LDP FRR 方式依赖 IP 网络中的 IP FRR 来实现。只有当 Liberal Label 的来源匹配存在的备份路由，即保留 Liberal Label 来自备份路由出接口和下一跳，并且满足备份 LSP 触发策略，同时没有根据该备份路由手工配置的备份 LSP 存在的时候，才能够为之建立备份 LSP 并下发转发表项。Auto LDP FRR 策略默认是 32 位的备份路由触发 LDP 建立备份 LSP。

在 Manual LDP FRR 和 Auto LDP FRR 同时满足创建条件的时候，优先建立手工配置的 LDP FRR。

4.4.2 LDP FRR 的实现原理

在自由标签保持方式下，LSR 可以从任何邻居 LSR 收到对于特定 FEC 的标签映射消息，但只有从该 FEC 对应当前有效路由下一跳发送来的标签映射会生成标签转发表，从而建立 LSP。通过 LDP FRR 也可以为来自非该 FEC 对应的当前有效路由的下一跳的标签映射生成 LSP，并作为主 LSP 的备份，建立转发表项，下发到转发表中，作为主转发表项的备份。当主 LSP 故障时，能快速切换到备份 LSP，避免流量的丢失。

如图 4-8 所示，LSR_1 到 LSR_2 的优选路由路径为 LSR_1-LSR_2，次优路由路径为 LSR_1-LSR_3-LSR_2。当 LSR_1 收到 LSR_3 发来的标签映射消息后，会和路由比较，因为 LSR_1 到 LSR_2 的当前有效路由的下一跳不是 LSR_3，所以 LSR_1 会把这个标签存为 Liberal Label。如果该 Liberal Label 的来源对应的备份路由（经 LSR_3 到达 LSR_2）存在，就可以为该 Liberal Label 申请一个转发表项资源，创建以备份 LSP 作为主 LSP 的备用转发表项，和主 LSP 一起下发到转发平面，这样主 LSP 就和这条备份 LSP 关联起来了。

接口感知接口故障、BFD 感知接口故障、BFD 感知主 LSP 不通等，都能触发 LDP FRR 切换。当 LDP FRR 切换后，流量根据备用转发表项切换到备份 LSP 上，至此 LDP FRR 生效。之后的变化过程是路由从 LSR_1-LSR_2 收敛到 LSR_1-LSR_3-LSR_2，在新的路径（原来的备份路径）上根据路由新建 LSP，再把原来的主 LSP 删除，流量按照 LSR_1-LSR_3-LSR_2 上新建的 LSP 进行转发。

图 4-8 所示是 LDP FRR 的典型应用场景，对这种三角形拓扑支持情况较好，但对图 4-9 所示的口字型拓扑不一定能够完全支持。

在图 4-9 中，如果 LSR_1 到 LSR_4 的最优路由路径是 LSR_1-LSR_2-LSR_4（不与其他路径负载分担），LSR_3 就会收到来自 LSR_1 的 Liberal 标签，并绑定 LDP FRR。当 LSR_3-LSR_4 之间的链路故障时，流量会切换到 LSR_3-LSR_1-LSR_2-LSR_4，不会形成环路。

如果 LSR_1 到 LSR_4 的路由是 LSR_1-LSR_2-LSR_4 和 LSR_1-LSR_3-LSR_4 负载分担，LSR_3 作为 LSR_1 的下游邻居，不一定会收到来自 LSR_1 的 Liberal 标签。并且，

即使 LSR_3 有了该 Liberal 标签，绑定了 LDP FRR，发生切换流量到达 LSR_1 后还很有可能会转发给 LSR_3，从而形成环路，直至 LSR_1 到 LSR_4 的路由收敛为 LSR_1-LSR_2-LSR_4。



图 4-8　LDP FRR 示例

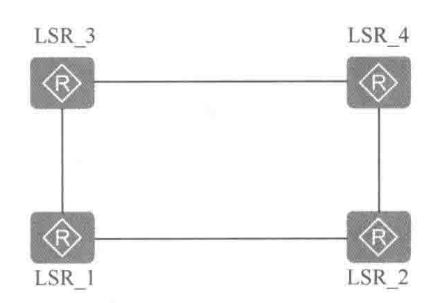


图 4-9　LDP FRR-口字型拓扑

4.4.3　配置 LDP FRR

4.4.1 节已介绍到，LDP FRR 功能的实现方式有两种：Manual（手动）LDP FRR、Auto（自动）LDP FRR。Manual LDP FRR 是采取手工配置方式来实现的，配置工程量比较大，适用于结构比较简单的网络。Auto LDP FRR 是通过 FRR 功能自动实现的，配置过程比较简单，适用于结构比较复杂的大型网络。可根据实际情况选择其中一种配置。

1. 配置 Manual LDP FRR

配置 Manual LDP FRR 的方法见表 4-9，**仅需要在 Ingress 或 Transit 节点上进行配置**，具体是在 Ingress 节点，还是在 Transit 节点上配置，要视备用 LDP LSP 的路径与主用 LDP LSP 的分支处是在哪个节点上。

注意：在 Manual LDP FRR 的配置中，备份 LSP 必须是 Liberal 状态的 LSP，即备份 LSP 的 Ingress 到 Egress 的路由状态必须是非活跃。

在配置 Manual LDP FRR 之前，需完成以下任务。

■ 配置本地 LDP 会话。

■ 如果配置基于 BFD 的 Manual LDP FRR，还需要完成 BFD 单跳检测的配置，参见《华为路由器学习指南》。

说明：因为在手动方案中是明确指定备份 LDP LSP 的路径，不需要 IGP FRR 参与，设备上不需要同时配置 IGP FRR。

表 4-9　Manual LDP FRR 的配置步骤

步骤	命令	说明
1	**system-view** 例如：<Huawei> **system-view**	进入系统视图
2	**interface** *interface-type interface-number* 例如：[Huawei] **interface** gigabitethernet 1/0/0	进入要使能 LDP FRR 功能的主用 LDP LSP 公网接口的接口视图

（续表）

<table>
<tr><th>步骤</th><th>命令</th><th>说明</th></tr>
<tr><td>3</td><td>mpls ldp frr nexthop nexthop-address [ip-prefix ip-prefix-name] [priority priority]
例如：[Huawei-GigabitEthernet 1/0/0] mpls ldp frr nexthop 10.1.1.2</td><td>在以上主用公网接口上使能 LDP FRR。命令中的参数说明如下。
• nexthop nexthop-address：指定备份 LSP 的下一跳 IP 地址。
• ip-prefix ip-prefix-name：可选参数，设置与指定 IP 前缀列表名称定义的 IP 前缀匹配的 FEC 才能触发生成备份 LSP，必须是已存在的 IP 前缀列表名称。
• priority priority：可选参数，指定备份 LSP 的优先级，整数形式，取值范围是 1～65535。缺省值是 50。优先级的值越大，该备份 LSP 的优先级越低。
【说明】执行此命令配置备份 LSP 的下一跳 IP 地址时。
• 在同一个接口下，最多可以有 10 项不同优先级的 LDP FRR 配置，但最终根据优先级只会生成一条备份 LSP。
• 可以在同一接口上配置多个不同的下一跳，即为主 LSP 配置多个不同出接口的备份 LSP。
• 也可以在同一接口下为相同下一跳配置不同的前缀列表。
➢ 如果不指定前缀列表，则 LDP FRR 会试图为本接口的所有 LSP 在 nexthop-address 参数所指定的路径上建立备份 LSP。
➢ 如果指定前缀列表中只有 DENY 项，则不允许该接口上被 DENY 的 FEC 对应的 LSP 在 nexthop-address 参数所指定的路径上建立备份 LSP。
➢ 如果指定前缀列表中只有 PERMIT 项，则只允许该接口上被 PERMIT 的 FEC 对应的 LSP 在 nexthop-address 参数所指定的路径上建立备份 LSP。如果指定前缀列表中既有 PERMIT 项又有 DENY 项，则只有 PERMIT 有效，即只允许该接口上被 PERMIT 的 FEC 对应的 LSP 在 nexthop-address 参数所指定的路径上建立备份 LSP。
LDP GR 期间禁止使能或去使能 LDP FRR 功能。混合应用 Manual LDP FRR 和 IP FRR 情况下，优先选择 IP FRR。使能 LDP 功能时，接口视图下的 LDP FRR 配置不会被自动删除，但 LDP FRR 功能已经失效。
缺省情况下，接口上没有使能 LDP FRR 功能，可用 undo mpls ldp frr [nexthop nexthop-address] [ip-prefix ip-prefix-name] [priority priority] 命令在接口上去使能 LDP FRR 功能</td></tr>
<tr><td colspan="3">以下步骤仅当配置基于静态 BFD 的 LDP FRR 时才需要执行</td></tr>
<tr><td>4</td><td>quit
例如：[Huawei-GigabitEthernet 1/0/0] quit</td><td>返回系统视图</td></tr>
<tr><td>5</td><td>bfd session-name
例如：[Huawei] bfd 4L3Int</td><td>进入已经创建的 BFD 会话视图。参数 session-name 必须是已在 4.2.2 节已创建的静态 BFD 的 LDP LSP 会话。
可用 undo bfd 命令删除指定的 BFD 会话</td></tr>
</table>

（续表）

步骤	命令	说明
6	**process-pst** 例如：[Huawei-bfd-session-4L3Int] **process-pst**	使能系统在 BFD 会话状态变化时修改端口状态表 PST（Port State Table）功能，允许 BFD 通告 LDP LSP。如果允许 BFD 修改端口状态表 PST，当检测到 BFD 会话状态变为 Down 时，系统将更改 PST 中相应表项。 缺省情况下，BFD 会话不使能通告联动检测业务，可用 **undo process-pst** 命令恢复缺省配置
7	**commit** 例如：[Huawei-bfd-session-4L3Int] **commit**	提交配置

2. 配置 Auto LDP FRR

配置 Auto LDP FRR 的方法如表 4-10 所示。**仅需在 Ingress 或 Transit 节点上进行配置**，具体要视备用 LDP LSP 的路径与主用 LDP LSP 的分支处是在哪个节点上。

在配置 Auto LDP FRR 之前，需完成以下任务。

- 配置本地 LDP 会话。
- 配置 Auto IGP FRR 功能，OSPF、IS-IS 协议均支持 Auto FRR 功能。

说明 因为在自动方案中仅是启动了 FRR 功能，不明确指定备份 LDP LSP 的路径，备份 LDP LSP 路径仍需要通过 IGP FRR 中的 LFA（Loop Free Alternate，无环路交替）算法进行自动计算得出的，故先需要在设备上配置 IGP FRR。

表 4-10　　Auto LDP FRR 的配置步骤

步骤	命令	说明
1	**system-view** 例如：<Huawei> **system-view**	进入系统视图
2	**mpls ldp** 例如：[Huawei] **mpls ldp**	进入 MPLS-LDP 视图
3	**auto-frr lsp-trigger** { **all** | **host** | **ip-prefix** *ip-prefix-name* | **none** } 例如：[Huawei-mpls-ldp] **auto-frr lsp-trigger host**	配置触发 LDP 建立备份 LSP 的策略。如果需要调整备份 LDP LSP 的建立策略，可以执行本命令。LDP GR 期间不允许修改备份 LSP 的触发策略。命令中的参数和选项说明如下。 • **all**：多选一选项，指定所有的备份路由都会触发 LDP 建立备份 LSP。 • **host**：多选一选项，指定 32 位地址的备份路由才会触发 LDP 建立备份 LSP。 • **ip-prefix** *ip-prefix-name*：多选一参数，设置根据指定 IP 地址前缀列表触发 LDP 建立备份 LSP。 • **none**：多选一选项，指定所有的备份路由都不触发 LDP 建立备份 LSP。 【说明】Auto LDP FRR 依赖 IGP 的自动重路由功能，所以需要先配置好 IGP 的 Auto FRR 功能，OSP Auto FRR 的配置方法如表 4-11 所示，IS-IS Auto FRR 的配置方法见表 4-12。

（续表）

步骤	命令	说明
3	**auto-frr lsp-trigger** { **all** | **host** | **ip-prefix** *ip-prefix-name* | **none** } 例如：[Huawei-mpls-ldp] **auto-frr lsp-trigger host**	还可通过 **lsp-trigger** { **all** | **host** | **ip-prefix** *ip-prefix-name* | **none** } 命令设置触发建立 LSP 的策略，缺省情况下，仅根据 32 位地址的主机 IP 路由（不包括接口的 32 位地址的主机 IP 路由）触发 LDP 建立 LSP。如果同时配置了本命令和以上 **lsp-trigger** 命令，则建立的备份 LSP 会同时满足 LDP 建立 LSP 的触发策略以及 LDP 建立备份 LSP 的触发策略。 缺省情况下，32 位地址的备份路由触发 LDP 建立备份 LSP，可用 **undo auto-frr lsp-trigger** 命令恢复缺省配置

表 4-11 OSPF Auto FRR 的配置步骤

步骤	命令	说明
1	**system-view** 例如：<Huawei> **system-view**	进入系统视图
2	**ospf** [*process-id* | **router-id** *router-id* | **vpn-instance** *vpn-instance-name*] * 例如：[Huawei] **ospf** 100	启动 OSPF 进程，进入 OSPF 视图。命令中的参数说明如下。 • *process-id*：可多选参数，指定要启动的 OSPF 进程号，整数形式，取值范围是 1～65535。缺省值是 1。 • **router-id** *router-id*：可多选参数，代表当前设备的 Router ID，点分十进制格式。 • **vpn-instance** *vpn-instance-name*：可多选参数，用于指定所启动的路由进程所属的 VPN 实例的名称
3	**frr** 例如：[Huawei-ospf-100] **frr**	进入 OSPF IP FRR 视图
4	**loop-free-alternate** 例如：[Huawei-ospf-100-frr] **loop-free-alternate**	使能 OSPF IP FRR 特性，生成无环的备份链路。 LFA 是实现 IP FRR 的一种方式，这样设备可以生成无环的备份链路，实现 IP FRR 的基本功能。如果网络中有承载重要业务的节点链路不能成为其他链路的备份链路，请在配置 OSPF IP FRR 功能前，在连接该节点设备的接口上配置 **ospf frr block** 命令。这样，FRR 计算时，就不会再把该接口所连接的链路计算成备份链路。 缺省情况下，不使能 OSPF IP FRR 功能，可用 **undo loop-free-alternate** 命令取消 OSPF IP FRR 功能
5	**frr-priority static low** 例如：[Huawei-ospf-100-frr] **frr-priority static low**	（可选）指定利用 LFA 算法计算备份下一跳和备份出接口，使动态备份路径的优先级高于静态备份路径的优先级。 【说明】OSPF 有两种方式可以获得备份路径。 • 静态备份路径：在系统视图或 VPN 实例视图下执行 **ip frr route-policy** *route-policy-name* 命令使能 IP FRR 功能后，需用 **apply backup-interface** *interface-type interface-number* 命令指定备份出接口和用 **apply backup-nexthop** { *ipv4-address* | **auto** } 命令指定备份下一跳。 • 动态备份路径：由 **loop-free-alternate** 命令使能 OSPF IP FRR 功能后，利用 LFA 算法计算备份下一跳和备份出接口。

（续表）

步骤	命令	说明
5	**frr-priority static low** 例如：[Huawei-ospf-100-frr] **frr-priority static low**	缺省情况下，静态备份路径的优先级高于动态备份路径的优先级，即静态备份路径会被优选，可用 **undo frr-priority static** 命令去使能该功能。但是，由于静态备份路径的灵活性较差，当备份路径出现故障时，静态备份路径不会自动更新，而动态备份路径可以自动更新。因此，为了保证备份路径的及时更新，可以配置本命令指定利用 LFA 算法计算备份下一跳和备份出接口，使动态备份路径的优先级高于静态备份路径的优先级
6	**frr-policy route route-policy** *route-policy-name* 例如：[Huawei-ospf-100-frr] **frr-policy route route-policy** abc	（可选）配置 OSPF IP FRR 过滤策略。参数 *route-policy-name* 用来指定 OSPF IP FRR 备份路由的过滤策略的名称。本命令是覆盖式命令，以最后一次配置为准。 配置了 OSPF IP FRR 过滤策略后，只有满足过滤条件的 OSPF 路由备份路由才能下发转发表。如果希望保护经过某条特定 OSPF 路由的流量时，可以通过设置过滤策略，使该 OSPF 路由满足过滤条件，则该 OSPF 路由的备份路由加入转发表中。当这条路由出现故障时，OSPF 可以快速将流量切换到备份路由上。 缺省情况下，不对使能 OSPF IP FRR 功能的备份路由进行过滤，可用 **undo frr-policy route** 命令取消 OSPF IP FRR 的备份路由的过滤功能

表 4-12　**IS-IS Auto FRR** 的配置步骤

步骤	命令	说明
1	**system-view** 例如：<Huawei> **system-view**	进入系统视图
2	**isis** [*process-id*] 例如：[Huawei] **isis** 1	使能 IS-IS 路由进程，进入 IS-IS 视图。
3	**frr** 例如：[Huawei-isis-1] **frr**	使能 FRR 并进入 IS-IS FRR 视图。 IS-IS Auto FRR 可以将流量快速切换到备份链路上，使流量中断的时间小于 50ms，从而达到保护流量的目的，因此极大地提高了 IS-IS 网络的可靠性。 缺省情况下，未使能 IS-IS FRR 功能，可用 **undo frr** 命令去使能 IS-IS FRR 功能
4	**frr-policy route route-policy** *route-policy-name* 例如：[Huawei-isis-1-frr] **frr-policy route route-policy** abc	（可选）利用过滤策略过滤备份路由，使只有通过过滤策略的备份路由才可以加入路由表。 用户可以根据需要，配置过滤策略，使得满足指定条件的 IS-IS 路由备份路由加入到 IP 路由表，并下发到转发表。当主路由发生故障时，系统可以快速将转发流量切换到 IS-IS 备份路由上，从而实现流量保护
5	**loop-free-alternate** [**level-1** | **level-2** | **level-1-2**] 例如：[Huawei-isis-1-frr] **loop-free-alternate**	使能 IS-IS Auto FRR 利用 LFA（Loop-free Alternate）算法计算无环备份路由。只有执行本命令之后，IS-IS 的 Auto FRR 功能才会生效。命令中的选项说明如下。 • **level-1**：多选一可选项，指定 Level-1 级别 IS-IS Auto FRR 并生成无环备份路由。如果不指定 Level，则在 Level-1 和 Level-2 上都使能 IS-IS Auto FRR 并生成备份路由。 • **level-2**：多选一可选项，指定 Level-2 级别 IS-IS Auto FRR 并生成无环备份路由。如果不指定 Level，则在 Level-1 和 Level-2 上都使能 IS-IS Auto FRR 并生成备份路由

（续表）

步骤	命令	说明
5	**loop-free-alternate** [**level-1** | **level-2** | **level-1-2**] 例如：[Huawei-isis-1-frr] **loop-free-alternate**	• **level-1-2**：多选一可选项，同时指定 Level-1 和 Level-2 级别 IS-IS Auto FRR 并生成无环备份路由 缺省情况下，未使能 IS-IS Auto FRR 利用 LFA（Loop-free Alternate）算法计算无环备份路由，可用 **undo loop-free-alternate** [**level-1** | **level-2** | **level-1-2**]命令用来去使能 IS-IS Auto FRR 利用 LFA 算法计算无环备份路由
6	**undo isis lfa-backup** [**level-1** | **level-2** | **level-1-2**] 例如：[Huawei-GigabitEthernet 1/0/0] **undo isis lfa-backup**	（可选）阻止接口参与 LFA（Loop Free Alternate）计算。在网络部署的过程中，为了便于流量管理，避免在主链路故障时流量转发路径的不确定性，可以阻止某些接口参与 LFA 计算，取消这些接口成为备份接口的能力。命令中的选项说明如下。 • **level-1**：多选一可选项，指定接口在 Level-1 范围内成为备份接口。 • **level-2**：多选一可选项，指定接口在 Level-2 范围内成为备份接口。 • **level-1-2**：多选一可选项，指定接口在 Level-1 和 Level-2 范围内成为备份接口。 如果不指定 Level，则接口在 Level-1 和 Level-2 上都使阻止参与 LFA 计算。 缺省情况下，使能 IS-IS 的接口可以参与 LFA 计算，可用 **undo isis lfa-backup** [**level-1** | **level-2** | **level-1-2**]命令去使能 IS-IS 接口参与 LFA 计算

为了实现毫秒级的快速切换，还需要同时配置静态 BFD 检测 LDP LSP 或者配置动态 BFD 检测 LDP LSP，分别参见 4.2.2 节和 4.2.3 节。配置好后可执行 **display mpls lsp** 命令查看使能了 LDP FRR 的 LSP 信息。

4.4.4 Manual LDP FRR 配置示例

如图 4-10 所示，网络拓扑结构不复杂并且稳定，部署了 MPLS LDP 业务。LSRA 到 LSRC 之间存在主备两条 LSP，其中 LSRA→LSRC 为主用 LSP，LSRA→LSRB→LSRC 为备份 LSP。主链路发生故障时，造成业务中断、流量丢失。要求在主用 LSP 发生故障的时候，流量能够快速切换到备份 LSP。

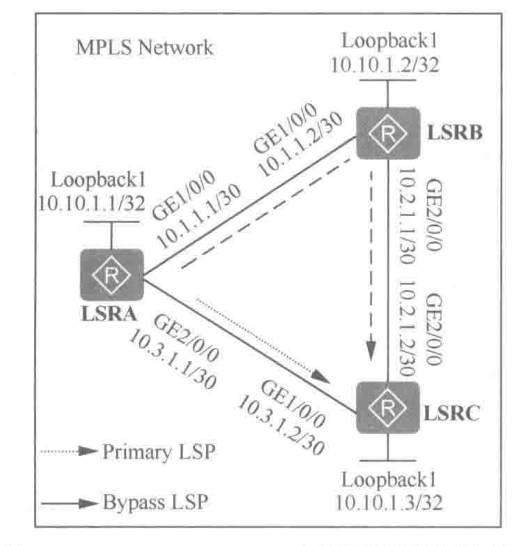


图 4-10 Manual LDP FRR 配置示例的拓扑结构

1. 基本配置思路分析

本示例其实既可以采用 Manual LDP FRR 方式，也可以采用 Auto LDP FRR，但本示例中的网络结构不复杂且稳定，所以采用 Manual LDP FRR 实现方式更为简单。

根据 4.4.3 节表 4-9 介绍可知，Manual LDP FRR 方式的配置比较简单，只需在路径分支的 Ingress 或 Transit 节点的主 LSP 出接

口上指定备份 LSP 的下一跳 IP 地址，或同时指定其允许触发建立备份 LSP 的路由策略、备份 LSP 的优先级。

本示例中因为只建立一条备份 LSP，不对触发建立备份 LSP 的流量进行过滤，且路径分支是在 Ingress 节点 LSRA 上，所以可以直接在 LSRA 主 LSP 的出接口 GE2/0/0 上配置 FRR 的备份 LSP 的下一跳 IP 地址即可。当然在此之前仍要完成整个骨干网的 MPLS 配置，实现各节点间的三层互通，并在相邻节点间建立本地 LDP 会话。

根据以上分析可得出本示例的基本配置思路如下。

（1）在骨干网各节点上配置各接口 IP 地址和 OSPF 协议，使骨干网各节点间三层互通。

（2）在骨干网各节点上配置基本 MPLS 能力和本地 LDP 会话。

（3）在主用 LDP LSP 的 LSRA 节点的 GE2/0/0 接口上使能 Manual LDP FRR，指定用于生成备用 LSP 的下一跳地址为 LSRB 的 GE1/0/0 接口 IP 地址。

说明 在 Manual LDP FRR 的组网中，备份 LSP 必须是 Liberal 状态的 LSP，即在使能 FRR 的节点上执行 **display ip routing-table** *ip-address* **verbose** 命令可以发现备份 LSP 的路由状态是“Inactive Adv”。

2. 具体配置步骤

（1）在骨干网各节点上配置各接口的 IP 地址和 OSPF 协议，实现整个骨干网三层互通。

\# LSRA 上的配置。

```
<Huawei> system-view
[Huawei] sysname LSRA
[LSRA] interface loopback 1
[LSRA-LoopBack1] ip address 10.10.1.1 32
[LSRA-LoopBack1] quit
[LSRA] interface gigabitethernet 1/0/0
[LSRA-GigabitEthernet1/0/0] ip address 10.1.1.1 30
[LSRA-GigabitEthernet1/0/0] quit
[LSRA] interface gigabitethernet 2/0/0
[LSRA-GigabitEthernet2/0/0] ip address 10.3.1.1 30
[LSRA-GigabitEthernet2/0/0] quit
[LSRA] ospf 1
[LSRA-ospf-1] area 0
[LSRA-ospf-1-area-0.0.0.0] network 10.10.1.1 0.0.0.0
[LSRA-ospf-1-area-0.0.0.0] network 10.1.1.0 0.0.0.3
[LSRA-ospf-1-area-0.0.0.0] network 10.3.1.0 0.0.0.3
[LSRA-ospf-1-area-0.0.0.0] quit
[LSRA-ospf-1] quit
```

\# LSRB 上的配置。

```
<Huawei> system-view
[Huawei] sysname LSRB
[LSRB] interface loopback 1
[LSRB-LoopBack1] ip address 10.10.1.2 32
[LSRB-LoopBack1] quit
[LSRB] interface gigabitethernet 1/0/0
[LSRB-GigabitEthernet1/0/0] ip address 10.1.1.2 30
[LSRB-GigabitEthernet1/0/0] quit
```

```
[LSRB] interface gigabitethernet 2/0/0
[LSRB-GigabitEthernet2/0/0] ip address 10.2.1.1 30
[LSRB-GigabitEthernet2/0/0] quit
[LSRB] ospf 1
[LSRB-ospf-1] area 0
[LSRB-ospf-1-area-0.0.0.0] network 10.10.1.2 0.0.0.0
[LSRB-ospf-1-area-0.0.0.0] network 10.1.1.0 0.0.0.3
[LSRB-ospf-1-area-0.0.0.0] network 10.2.1.0 0.0.0.3
[LSRB-ospf-1-area-0.0.0.0] quit
[LSRB-ospf-1] quit
```

LSRC 上的配置。

```
<Huawei> system-view
[Huawei] sysname LSRC
[LSRC] interface loopback 1
[LSRC-LoopBack1] ip address 10.10.1.3 32
[LSRC-LoopBack1] quit
[LSRC] interface gigabitethernet 1/0/0
[LSRC-GigabitEthernet1/0/0] ip address 10.3.1.2 30
[LSRC-GigabitEthernet1/0/0] quit
[LSRC] interface gigabitethernet 2/0/0
[LSRC-GigabitEthernet2/0/0] ip address 10.2.1.2 30
[LSRC-GigabitEthernet2/0/0] quit
[LSRC] ospf 1
[LSRC-ospf-1] area 0
[LSRC-ospf-1-area-0.0.0.0] network 10.10.1.3 0.0.0.0
[LSRC-ospf-1-area-0.0.0.0] network 10.3.1.0 0.0.0.3
[LSRC-ospf-1-area-0.0.0.0] network 10.2.1.0 0.0.0.3
[LSRC-ospf-1-area-0.0.0.0] quit
[LSRC-ospf-1] quit
```

以上配置完成后，在各节点上执行 **display ip routing-table** 命令，可以看到相互之间都学到了彼此的路由。以下是在 LSRA 上执行该命令的输出示例，发现已学习到了 LSRB 和 LSRC 的 Loopback 接口主机路由（参见输出信息中的粗体字部分）。

```
[LSRA] display ip routing-table
 Route Flags: R - relay, D - download to fib
 ------------------------------------------------------------------------------
 Routing Tables: Public
          Destinations : 14        Routes : 15
 Destination/Mask    Proto  Pre  Cost      Flags NextHop         Interface
        10.10.1.1/32  Direct 0    0           D   127.0.0.1       LoopBack1
        10.10.1.2/32  OSPF   10   1           D   10.1.1.2        GigabitEthernet1/0/0
        10.10.1.3/32  OSPF   10   1           D   10.3.1.2        GigabitEthernet2/0/0
         10.1.1.0/30  Direct 0    0           D   10.1.1.1        GigabitEthernet1/0/0
         10.1.1.1/32  Direct 0    0           D   127.0.0.1       GigabitEthernet1/0/0
       10.1.1.255/32  Direct 0    0           D   127.0.0.1       GigabitEthernet1/0/0
         10.2.1.0/30  OSPF   10   2           D   10.3.1.2        GigabitEthernet2/0/0
                      OSPF   10   2           D   10.1.1.2        GigabitEthernet1/0/0
         10.3.1.0/30  Direct 0    0           D   10.3.1.1        GigabitEthernet2/0/0
         10.3.1.1/32  Direct 0    0           D   127.0.0.1       GigabitEthernet2/0/0
       10.3.1.255/32  Direct 0    0           D   127.0.0.1       GigabitEthernet2/0/0
          127.0.0.0/8 Direct 0    0           D   127.0.0.1       InLoopBack0
        127.0.0.1/32  Direct 0    0           D   127.0.0.1       InLoopBack0
 127.255.255.255/32   Direct 0    0           D   127.0.0.1       InLoopBack0
 255.255.255.255/32   Direct 0    0           D   127.0.0.1       InLoopBack0
```

（2）在骨干网各节点上配置基本的 MPLS 能力和 LDP 本地会话，通过 LDP 自动协商建立双向 LDP LSP。

#　LSRA 上的配置。

```
[LSRA] mpls lsr-id 10.10.1.1
[LSRA] mpls
[LSRA-mpls] quit
[LSRA] mpls ldp
[LSRA-mpls-ldp] quit
[LSRA] interface gigabitethernet 1/0/0
[LSRA-GigabitEthernet1/0/0] mpls
[LSRA-GigabitEthernet1/0/0] mpls ldp
[LSRA-GigabitEthernet1/0/0] quit
[LSRA] interface gigabitethernet 2/0/0
[LSRA-GigabitEthernet2/0/0] mpls
[LSRA-GigabitEthernet2/0/0] mpls ldp
[LSRA-GigabitEthernet2/0/0] quit
```

#　LSRB 上的配置。

```
[LSRB] mpls lsr-id 10.10.1.2
[LSRB] mpls
[LSRB-mpls] quit
[LSRB] mpls ldp
[LSRB-mpls-ldp] quit
[LSRB] interface gigabitethernet 1/0/0
[LSRB-GigabitEthernet1/0/0] mpls
[LSRB-GigabitEthernet1/0/0] mpls ldp
[LSRB-GigabitEthernet1/0/0] quit
[LSRB] interface gigabitethernet 2/0/0
[LSRB-GigabitEthernet2/0/0] mpls
[LSRB-GigabitEthernet2/0/0] mpls ldp
[LSRB-GigabitEthernet2/0/0] quit
```

#　LSRC 上的配置。

```
[LSRC] mpls lsr-id 10.10.1.3
[LSRC] mpls
[LSRC-mpls] quit
[LSRC] mpls ldp
[LSRC-mpls-ldp] quit
[LSRC] interface gigabitethernet 1/0/0
[LSRC-GigabitEthernet1/0/0] mpls
[LSRC-GigabitEthernet1/0/0] mpls ldp
[LSRC-GigabitEthernet1/0/0] quit
[LSRC] interface gigabitethernet 2/0/0
[LSRC-GigabitEthernet2/0/0] mpls
[LSRC-GigabitEthernet2/0/0] mpls ldp
[LSRC-GigabitEthernet2/0/0] quit
```

上述配置完成后，相邻节点之间应该建立起 LDP 会话。在各节点上执行 **display mpls ldp session** 命令可以看到显示结果中 Status 项为“Operational”。以下是在 LSRA 上执行该命令的输出示例，从中可以看出已与 LSRB 和 LSRC 建立了 LDP 会话（参见输出信息中的粗体字部分）。

```
[LSRA] display mpls ldp session

 LDP Session(s) in Public Network
```

```
Codes: LAM(Label Advertisement Mode), SsnAge Unit(DDDD:HH:MM)
A '*' before a session means the session is being deleted.
------------------------------------------------------------------------------
PeerID              Status          LAM   SsnRole   SsnAge          KASent/Rcv
------------------------------------------------------------------------------
10.10.1.2:0         Operational DU       Passive    0000:00:01    8/8
10.10.1.3:0         Operational DU       Passive    0000:00:01    6/6
------------------------------------------------------------------------------
TOTAL: 2 session(s) Found.
```

（3）在 LSRA 中主 LSP 出接口 GE2/0/0 下使能 Manual LDP FRR，并指定用于生成备份 LSP 的下一跳地址为 LSRB 的 GE1/0/0 接口 IP 地址。

```
[LSRA] interface gigabitethernet 2/0/0
[LSRA-GigabitEthernet2/0/0] mpls ldp frr nexthop 10.1.1.2
[LSRA-GigabitEthernet2/0/0] quit
```

说明 如果还要使 LSRC 到 LSRA 之间的通信也具有 LDP FRR 功能，则还需要在 LSRC 主 LDP LSP 的出接口 GE1/0/0 上配置备份 LSP 的下一跳地址为 LSRB 的 GE2/0/0 接口 IP 地址。

3．配置结果验证

上述配置完成后，在 LSRA 节点上执行 **display mpls lsp** 命令，可以看到到达 LSRC 的 LSP 上存在 LDP FRR（参见输出信息中的粗体字部分）。

```
[LSRA] display mpls lsp
------------------------------------------------------------------------------
                    LSP Information: LDP LSP
------------------------------------------------------------------------------
FEC                 In/Out Label    In/Out IF                  Vrf Name
10.10.1.2/32        NULL/3            -/GE1/0/0
10.10.1.2/32        1024/3          -/GE1/0/0
10.10.1.3/32        NULL/3            -/GE2/0/0
   **LDP FRR**              /1025            /GE1/0/0
10.10.1.3/32        1025/3          -/GE2/0/0
   **LDP FRR**              /1025            /GE1/0/0
10.10.1.1/32        3/NULL            -/-
```

4.4.5 LDP Auto FRR 配置示例

如图 4-11 所示，网络结构拓扑复杂并且不稳定，部署了 MPLS LDP 业务。LSRA 到 LSRC 之间存在主备 LSP，其中 LSRA→LSRC 为主用 LSP，LSRA→LSRB→LSRC 为备份 LSP。主链路发生故障时，造成业务中断、流量丢失。要求在主用 LSP 发生故障的时候，流量能够快速切换到备份 LSP。

1．基本配置思路分析

本示例与 4.4.4 节所介绍的示例的主要区别就在于本示例的网络拓扑结构复杂且不稳定，所以不能采用 4.4.4 节介绍的 Manual LDP FRR 方式，而要采用 LDP Auto FRR 方式。

根据 4.4.3 节介绍的 Auto LDP FRR 配置方法可知，Auto LDP FRR 方式的配置也不复杂，主要是配置触发 LDP 建立备份 LSP 的策略，但在此之前还要根据骨干网所运行的 IGP 协议类型，使能 Auto IP FRR。有关 OSP 和 IS-IS 协议中的 Auto IP FRR 功能的配

置方法参见 4.4.3 节中的表 4-11 和表 4-12。当然，首先也是需要配置好骨干网的 MPLS 基本能力和各节点间的 LDP 本地会话。

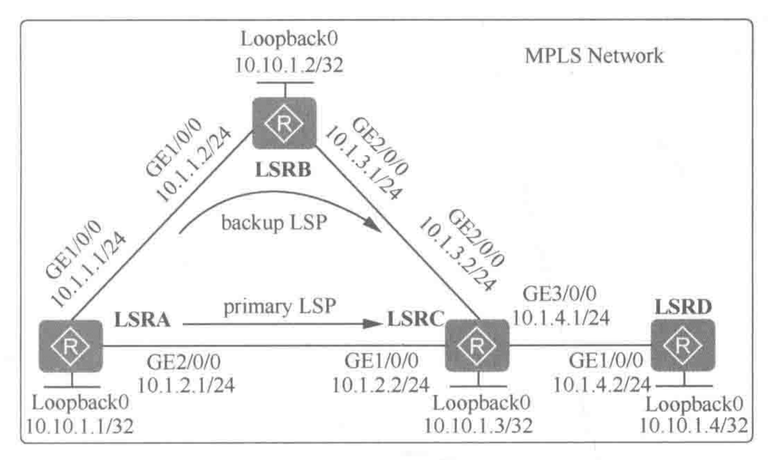


图 4-11　LDP Auto FRR 配置示例的拓扑结构

根据以上分析可得出本示例的基本配置思路如下。

（1）在骨干网各节点上配置各接口 IP 地址和 IS-IS 协议，使骨干网各节点间三层互通。

（2）在骨干网各节点上配置基本 MPLS 能力和本地 LDP 会话。

（3）在备份 LDP LSP 的 Ingress 节点 LSRA 上使有 IS-IS Auto FRR。

（4）在备份 LDP LSP 的 Ingress 节点 LSRA 上配置 LDP Auto FRR，使它自动计算备份 LSP 的下一跳和出接口。

2. 具体配置步骤

（1）在骨干网各节点上配置各接口 IP 地址和 IS-IS 协议。各节点设备均位于区域 16（对应十六进制为 0010）中，LSRA、LSRB、LSRC 和 LSRD 的系统 ID 分别为 1、2、3、4（对应十六进制为分别 0001、0002、0003、0004）。

LSRA 上的配置。

```
<Huawei> system-view
[Huawei] sysname LSRA
[LSRA] interface loopback 0
[LSRA-LoopBack0] ip address 10.10.1.1 32
[LSRA-LoopBack0] quit
[LSRA] interface gigabitethernet 1/0/0
[LSRA-GigabitEthernet1/0/0] ip address 10.1.1.1 24
[LSRA-GigabitEthernet1/0/0] quit
[LSRA] interface gigabitethernet 2/0/0
[LSRA-GigabitEthernet2/0/0] ip address 10.1.2.1 24
[LSRA-GigabitEthernet2/0/0] quit
[LSRA] isis 1
[LSRA-isis-1] network-entity 10.0000.0000.0001.00  #---指定网络实体名称为 10.0000.0000.0001.00，其中区域 ID 为 16，系统 ID 为 1
[LSRA-isis-1] quit
[LSRA] interface gigabitethernet 1/0/0
[LSRA-GigabitEthernet1/0/0] isis enable 1
[LSRA-GigabitEthernet1/0/0] quit
```

```
[LSRA] interface gigabitethernet 2/0/0
[LSRA-GigabitEthernet2/0/0] isis enable 1
[LSRA-GigabitEthernet2/0/0] quit
[LSRA] interface loopback 0
[LSRA-LoopBack0] isis enable 1
[LSRA-LoopBack0] quit
```

\# LSRB 上的配置。

```
<Huawei> system-view
[Huawei] sysname LSRB
[LSRB] interface loopback 0
[LSRB-LoopBack0] ip address 10.10.1.2 32
[LSRB-LoopBack0] quit
[LSRB] interface gigabitethernet 1/0/0
[LSRB-GigabitEthernet1/0/0] ip address 10.1.1.2 24
[LSRB-GigabitEthernet1/0/0] quit
[LSRB] interface gigabitethernet 2/0/0
[LSRB-GigabitEthernet2/0/0] ip address 10.1.3.1 24
[LSRB-GigabitEthernet2/0/0] quit
[LSRB] isis 1
[LSRB-isis-1] network-entity 10.0000.0000.0002.00
[LSRB-isis-1] quit
[LSRB] interface gigabitethernet 1/0/0
[LSRB-GigabitEthernet1/0/0] isis enable 1
[LSRB-GigabitEthernet1/0/0] quit
[LSRB] interface gigabitethernet 2/0/0
[LSRB-GigabitEthernet2/0/0] isis enable 1
[LSRB-GigabitEthernet2/0/0] quit
[LSRB] interface loopback 0
[LSRB-LoopBack0] isis enable 1
[LSRB-LoopBack0] quit
```

\# LSRC 上的配置。

```
<Huawei> system-view
[Huawei] sysname LSRC
[LSRC] interface loopback 0
[LSRC-LoopBack0] ip address 10.10.1.3 32
[LSRC-LoopBack0] quit
[LSRC] interface gigabitethernet 1/0/0
[LSRC-GigabitEthernet1/0/0] ip address 10.1.2.2 24
[LSRC-GigabitEthernet1/0/0] quit
[LSRC] interface gigabitethernet 2/0/0
[LSRC-GigabitEthernet2/0/0] ip address 10.1.3.2 24
[LSRC-GigabitEthernet2/0/0] quit
[LSRC] interface gigabitethernet 3/0/0
[LSRC-GigabitEthernet3/0/0] ip address 10.1.4.1 24
[LSRC-GigabitEthernet3/0/0] quit
[LSRC] isis 1
[LSRC-isis-1] network-entity 10.0000.0000.0003.00
[LSRC-isis-1] quit
[LSRC] interface gigabitethernet 1/0/0
[LSRC-GigabitEthernet1/0/0] isis enable 1
[LSRC-GigabitEthernet1/0/0] quit
[LSRC] interface gigabitethernet 2/0/0
[LSRC-GigabitEthernet2/0/0] isis enable 1
[LSRC-GigabitEthernet2/0/0] quit
```

```
[LSRC] interface gigabitethernet 3/0/0
[LSRC-GigabitEthernet3/0/0] isis enable 1
[LSRC-GigabitEthernet3/0/0] quit
[LSRC] interface loopback 0
[LSRC-LoopBack0] isis enable 1
[LSRC-LoopBack0] quit
```

#　LSRD 上的配置。

```
<Huawei> system-view
[Huawei] sysname LSRD
[LSRD] interface loopback 0
[LSRD-LoopBack0] ip address 10.10.1.4 32
[LSRD-LoopBack0] quit
[LSRD] interface gigabitethernet 1/0/0
[LSRD-GigabitEthernet1/0/0] ip address 10.1.4.2 24
[LSRD-GigabitEthernet1/0/0] quit
[LSRD] isis 1
[LSRD-isis-1] network-entity 10.0000.0000.0004.00
[LSRD-isis-1] quit
[LSRD] interface gigabitethernet 1/0/0
[LSRD-GigabitEthernet1/0/0] isis enable 1
[LSRD-GigabitEthernet1/0/0] quit
[LSRD] interface loopback 0
[LSRD-LoopBack0] isis enable 1
[LSRD-LoopBack0] quit
```

以上配置完成后，通过 **display ip routing-table** 命令可以查看到各节点间已相互学习到路由了。

（2）在骨干网各节点上配置基本 MPLS 能力和本地 LDP 会话。

#　LSRA 上的配置。

```
[LSRA] mpls lsr-id 10.10.1.1
[LSRA] mpls
[LSRA-mpls] quit
[LSRA] mpls ldp
[LSRA-mpls-ldp] quit
[LSRA] interface gigabitethernet 1/0/0
[LSRA-GigabitEthernet1/0/0] mpls
[LSRA-GigabitEthernet1/0/0] mpls ldp
[LSRA-GigabitEthernet1/0/0] quit
[LSRA] interface gigabitethernet 2/0/0
[LSRA-GigabitEthernet2/0/0] mpls
[LSRA-GigabitEthernet2/0/0] mpls ldp
[LSRA-GigabitEthernet2/0/0] quit
```

配置 LSRB。

```
[LSRB] mpls lsr-id 10.10.1.2
[LSRB] mpls
[LSRB-mpls] quit
[LSRB] mpls ldp
[LSRB-mpls-ldp] quit
[LSRB] interface gigabitethernet 1/0/0
[LSRB-GigabitEthernet1/0/0] mpls
[LSRB-GigabitEthernet1/0/0] mpls ldp
[LSRB-GigabitEthernet1/0/0] quit
[LSRB] interface gigabitethernet 2/0/0
```

```
[LSRB-GigabitEthernet2/0/0] mpls
[LSRB-GigabitEthernet2/0/0] mpls ldp
[LSRB-GigabitEthernet2/0/0] quit
```

配置 LSRC。

```
[LSRC] mpls lsr-id 10.10.1.3
[LSRC] mpls
[LSRC-mpls] quit
[LSRC] mpls ldp
[LSRC-mpls-ldp] quit
[LSRC] interface gigabitethernet 1/0/0
[LSRC-GigabitEthernet1/0/0] mpls
[LSRC-GigabitEthernet1/0/0] mpls ldp
[LSRC-GigabitEthernet1/0/0] quit
[LSRC] interface gigabitethernet 2/0/0
[LSRC-GigabitEthernet2/0/0] mpls
[LSRC-GigabitEthernet2/0/0] mpls ldp
[LSRC-GigabitEthernet2/0/0] quit
[LSRC] interface gigabitethernet 3/0/0
[LSRC-GigabitEthernet3/0/0] mpls
[LSRC-GigabitEthernet3/0/0] mpls ldp
[LSRC-GigabitEthernet3/0/0] quit
```

配置 LSRD。

```
[LSRD] mpls lsr-id 10.10.1.4
[LSRD] mpls
[LSRD-mpls] quit
[LSRD] mpls ldp
[LSRD-mpls-ldp] quit
[LSRD] interface gigabitethernet 1/0/0
[LSRD-GigabitEthernet1/0/0] mpls
[LSRD-GigabitEthernet1/0/0] mpls ldp
[LSRD-GigabitEthernet1/0/0] quit
```

以上配置完成后，在 LSRA 上执行 **display mpls lsp** 命令，查看已经建立的 LSP，发现 LSRA 到达 LSRB、LSRC 和 LSRD 的 LDP LSP 均已建立好。

```
[LSRA] display mpls lsp
-------------------------------------------------------------------------------
                     LSP Information: LDP LSP
-------------------------------------------------------------------------------
FEC               In/Out Label   In/Out IF                        Vrf Name
10.10.1.2/32      NULL/3          -/GE1/0/0
10.10.1.2/32      1024/3         -/GE1/0/0
10.10.1.3/32      NULL/3          -/GE2/0/0
10.10.1.3/32      1025/3         -/GE2/0/0
10.10.1.4/32      NULL/1026       -/GE2/0/0
10.10.1.4/32      1026/1026      -/GE2/0/0
10.10.1.1/32      3/NULL          -/-
```

说明 缺省情况下，仅 32 位掩码的主机路由可以建立 LDP LSP，所以骨干网中公网接口的直连路由是不会建立 LDP LSP 的，当然也可以通过 **lsp-trigger** { **all** | **host** | **ip-prefix** *ip-prefix-name* | **none** }命令设置触发建立 LDP LSP 的路由策略，使非 32 位掩码主机路由也可以建立 LDP LSP。

（3）在 LSRA 上使能 IS-IS Auto FRR，不过滤进入 IP 路由表的备份路由。

```
[LSRA] isis
[LSRA-isis-1] frr   #---使能 IP FRR
[LSRA-isis-1-frr] loop-free-alternate   #---使能 IS-IS Auto FRR 利用 LFA 算法计算无环备份路由
[LSRA-isis-1-frr] quit
[LSRA-isis-1] quit
```

以上配置完成后，执行 **display ip routing-table** 10.1.4.0 **verbose** 命令查看 LSRA 到 LSRC 和 LSRD 之间直连链路的路由信息。从显示信息可以看到，由于使能了 IS-IS Auto FRR 利用 LFA 算法计算无环备份路由，使从 LSRA 经 LSRB 到 LSRC 的备份路由也下发到 IP 路由表中了（参见以下输出信息中的粗体字部分）。

```
[LSRA] display ip routing-table 10.1.4.0 verbose
Route Flags: R - relay, D - download to fib
------------------------------------------------------------------------------
Routing Table : Public
Summary Count : 1

Destination: 10.1.4.0/24
     Protocol: ISIS-L1          Process ID: 1
   Preference: 15                     Cost: 20
      NextHop: 10.1.2.2          Neighbour: 0.0.0.0
        State: Active Adv              Age: 00h05m38s
          Tag: 0                  Priority: low
        Label: NULL                 QoSInfo: 0x0
   IndirectID: 0x0
 RelayNextHop: 0.0.0.0           Interface: GigabitEthernet2/0/0
     TunnelID: 0x0                   Flags: D
    BkNextHop: 10.1.1.2        BkInterface: GigabitEthernet1/0/0
      BkLabel: NULL            SecTunnelID: 0x0
 BkPETunnelID: 0x0          BkPESecTunnelID: 0x0
 BkIndirectID: 0x0
```

在备份路由也下发到 IP 路由表后，就可以在这条备份路由路径上建立到达 LSRB 和 LSRC 的备份 LDP LSP 了。所以此时，再在 LSRA 上执行 **display mpls lsp** 命令，就可以看到，在缺省情况下可以建立 LDP LSP 的 32 位主机路由下均触发建立了备份 LDP LSP。

```
[LSRA] display mpls lsp
------------------------------------------------------------------------------
                 LSP Information: LDP LSP
------------------------------------------------------------------------------
FEC                In/Out Label  In/Out IF              Vrf Name
10.10.1.2/32       NULL/3        -/GE1/0/0
   **LDP FRR**         /1025         /GE2/0/0
10.10.1.2/32       1024/3        -/GE1/0/0
   **LDP FRR**         /1025         /GE2/0/0
10.10.1.3/32       NULL/3        -/GE2/0/0
   **LDP FRR**         /1025         /GE1/0/0
10.10.1.3/32       1025/3        -/GE2/0/0
   **LDP FRR**         /1025         /GE1/0/0
10.10.1.4/32       NULL/1026     -/GE2/0/0
   **LDP FRR**         /1026         /GE1/0/0
10.10.1.4/32       1026/1026     -/GE2/0/0
   **LDP FRR**         /1026         /GE1/0/0
10.10.1.1/32       3/NULL        -/-
```

如果要使LSRA到达10.1.4.0/24网段的通信也可以采用备份LDP LSP，则可在LSRC上执行 lsp-trigger all 命令，改变 LSP 触发策略，使所有路由触发 LDP 建立 LSP。

```
[LSRC] mpls
[LSRC-mpls] lsp-trigger all
[LSRC-mpls] quit
```

此时，再在 LSRA 上执行 **display mpls lsp** 命令，查看已经建立的 LSP。从显示信息可以看到，24 位掩码的路由触发也建立了 LSP，参见输出信息中的粗体字部分。

```
[LSRA] display mpls lsp
-------------------------------------------------------------------------------
                 LSP Information: LDP LSP
-------------------------------------------------------------------------------
FEC                In/Out Label    In/Out IF                      Vrf Name
10.10.1.2/32       NULL/3          -/GE1/0/0
   **LDP FRR**          /1025        /GE2/0/0
10.10.1.2/32       1024/3         -/GE1/0/0
   **LDP FRR**          /1025        /GE2/0/0
10.10.1.3/32       NULL/3          -/GE2/0/0
   **LDP FRR**          /1025        /GE1/0/0
10.10.1.3/32       1025/3         -/GE2/0/0
   **LDP FRR**          /1025        /GE1/0/0
10.10.1.4/32       NULL/1026       -/GE2/0/0
   **LDP FRR**          /1026        /GE1/0/0
10.10.1.4/32       1026/1026      -/GE2/0/0
   **LDP FRR**          /1026        /GE1/0/0
10.10.1.1/32       3/NULL          -/-
10.1.1.0/24      1027/3       -/GE2/0/0
10.1.2.0/24      1028/3       -/GE2/0/0
```

（4）在 LSRA 上配置备份 LSP 的触发策略，使所有备份路由（包括允许触发 LDP LSP 的 24 位掩码路由）触发 LDP 建立备份 LSP。

```
[LSRA] mpls ldp
[LSRA-mpls-ldp] auto-frr lsp-trigger all
[LSRA-mpls-ldp] quit
```

3. 配置结果验证

上述配置完成后，在 LSRA 节点上执行 **display mpls lsp** 命令，查看备份 LSP 的建立情况。从显示信息可以看到，主 LSP 路径上 24 位掩码的路由也触发建立了 LSP，参见以上输出信息中的粗体字部分。这样，当主 LDP LSP 出现故障时，不管是从 LSRA 到 LSRC，还是到达 10.1.4.0/24 网段的通信都可以选择备份 LDP LSP 进行通信。

```
[LSRA] display mpls lsp
-------------------------------------------------------------------------------
                 LSP Information: LDP LSP
-------------------------------------------------------------------------------
FEC                In/Out Label    In/Out IF                      Vrf Name
10.10.1.2/32       NULL/3          -/GE1/0/0
   **LDP FRR**          /1025        /GE2/0/0
10.10.1.2/32       1024/3         -/GE1/0/0
   **LDP FRR**          /1025        /GE2/0/0
10.10.1.3/32       NULL/3          -/GE2/0/0
   **LDP FRR**          /1025        /GE1/0/0
10.10.1.3/32       1025/3         -/GE2/0/0
   **LDP FRR**          /1025        /GE1/0/0
```

```
10.10.1.4/32          NULL/1026        -/GE2/0/0
    **LDP FRR**             /1026          /GE1/0/0
10.10.1.4/32          1026/1026        -/GE2/0/0
    **LDP FRR**             /1026          /GE1/0/0
10.10.1.1/32          3/NULL            -/-
10.1.1.0/24           1027/3           -/GE2/0/0
10.1.2.0/24           1028/3           -/GE2/0/0
    **LDP FRR**        /1027          /GE1/0/0
```

4.5　LDP GR 配置与管理

LDP GR（Graceful Restart）利用 MPLS 转发平面与控制平面分离的特点，实现设备在协议重启或主备倒换（将备用主控板倒换为主用主控板）时转发不中断。其实这也是由 IP 网络中的 GR 功能扩展而来的。IP 网络中各种路由协议都有 GR 功能，可以实现在路由协议重启或主备倒换时转发不中断。

说明　通过主备倒换功能，可以将备用主控板倒换为主用主控板，实现主用主控板和备用主控板之间的冗余备份。执行主备倒换后，设备运行的主用主控板将重新启动，且启动后成为备用主控板；设备正在运行的备用主控板将成为主用主控板。

4.5.1　LDP GR 工作原理

在 MPLS 网络中，设备协议重启或主备倒换时，设备会删除转发平面上原有的标签转发表项，导致数据转发中断。通过 LDP GR 可以解决此问题，提高网络的可靠性。因为 LDP GR 可在设备协议重启或主备倒换时，利用控制平面和转发平面分离的特点，保留原来的标签转发表项，这样设备依然可以根据原来的标签转发表项转发报文，从而保证数据传输不会中断。同时，在协议重启或主备倒换后，设备还可在邻居设备的协助下恢复到重启之前的状态。

LDP GR 是基于 NSF（None Stop Forwarding，不间断转发）理念设计的一种高可靠性技术。在 GR 的过程中需要有 GR Restarter 和 GR Helper 两种角色的设备的参与。

■ GR Restarter：具备 GR 能力，指由管理员手工触发或控制平面异常而重启协议的设备，即要进行 GR 的设备。

■ GR Helper：也具备 GR 能力，与重启的 GR Restarter 保持邻居关系，并协助其恢复重启前的转发状态。

说明　仅 AR3260-S 设备支持作为 GR Restarter 和 GR Helper，其他设备只支持作为 GR Helper。使能或去使能 LDP GR 功能、修改 LDP GR 相关定时器的值都会导致 LDP 会话重建。

在整个 LDP GR 过程中涉及以下三个定时器。

■ 转发状态保持定时器：Forwarding State Holding Timer，也称邻居存活定时器，标识了 LDP GR 过程持续的时间。

■ 重连接定时器：Reconnect Timer：GR Restarter 发生协议重启或主备倒换后，GR Helper 检测到和 GR Restarter 的 LDP 会话失败，将启动重连接定时器，等待 LDP 会话的重新建立。

■ 恢复定时器：Recovery Timer：LDP 会话重新建立后，GR Helper 启动恢复定时器，等待 LSP 的恢复。

以上三个 LDP GR 相关的定时器的具体说明如表 4-13 所示。

表 4-13 LDP GR 相关的三个定时器说明

定时器	描述	使用建议
Neighbor-liveness 定时器（邻居存活定时器）	邻居存活定时器的值标识了 LDP GR 持续的时间。 GR Restarter 配置的 Neighbor-liveness 定时器的值就是 Forwarding State Holding 定时器（转发状态保持定时器）的值	当网络中 LSP 的数量较少时，可以配置较小的邻居存活定时器的值，短时间内结束 GR
Reconnect 定时器（LDP 会话重连接定时器）	GR Restarter 发生主备倒换后，GR Helper 检测到和 GR Restarter 的 LDP 会话失败，将启动 Reconnect 定时器，等待 LDP 会话的重新建立。 GR Helper 实际生效的 Reconnect 定时器的值是 GR Helper 配置的 Neighbor-liveness 定时器的值和 GR Restarter 配置的 Reconnect 定时器的值中的较小值	
Recovery 定时器（LSP 恢复定时器）	LDP 会话重新建立后，GR Helper 启动 Recovery 定时器，等待 LSP 的恢复。 GR Helper 实际生效的 Recovery 定时器的值是 GR Helper 配置的 Recovery 定时器的值和 GR Restarter 配置的 Recovery 定时器的值中的较小值	存在大量路由的网络中，网络故障时，为了防止缺省 300s 内无法恢复所有 LSP，可以调大 LSP 恢复定时器的值

LDP GR 的具体实现流程如图 4-12 所示，具体描述如下。

（1）首先，GR Restarter 和 GR Helper 之间要建立 LDP 会话。在 LDP 会话建立过程中，两者要协商 GR 能力。双方在发送的 Initialization（初始化）消息中携带的 FT（Fault Tolerance，容错）标记位为 1（标识状态为 on），标识它们支持 LDP GR。

（2）当 GR Restarter 协议重启或主备倒换时，会启动 MPLS 转发状态保持定时器，保留当前标签转发表项，并将标签转发表项置为 Stale（陈旧）状态，然后对 GR Helper 发送 LDP 初始化消息。GR Helper 发现与 GR Restarter 之间的 LDP 会话失败后，将保留与 GR Restarter 相关的标签转发表项，也将通过该 LDP 会话接收的 FEC 标签映射置为 Stale 状态，并启动重连接定时器。

（3）GR Restarter 协议重启或主备倒换后，重新建立与 GR Helper 的 LDP 会话。如果在 GR Helper 启动的重连定时器超时前没有成功与 GR Restarter 建立 LDP 会话，则 GR Helper 删除 GR Restarter 相关、标记为 Stale 的 FEC 标签映射及对应的标签转发表项。如果在重连定时器超时前，重新成功建立 LDP 会话，则 GR Restarter 将转发状态保持定时器的剩余时间作为恢复定时器时间值通告给 GR Helper。

（4）GR Restarter 和 GR Helper 之间重新建立 LDP 会话后，GR Helper 启动 LSP 恢复定时器。在恢复定时器超时前，GR Restarter 和 GR Helper 在新建立的 LDP 会话上交互标签映射消息。GR Helper 协助 GR Restarter 恢复转发表项，同时 GR Restarter 也会协助 GR Helper 恢复转发表项。

GR Restarter 接收到标签映射后，与本地标签转发表进行比较：如果标签转发表中存在与标签映射一致的表项，则删除该表项的 Stale 标记；否则，按照正常的 LDP 处理流程，添加新的标签转发表项。GR Helper 接收到标签映射后，也与本地保存的 FEC 标签映射进行比较：如果存在一致的标签映射，则删除该 FEC 标签映射的 Stale 标记；否则，按照正常的 LDP 处理流程，添加新的 FEC 标签映射及对应的标签转发表项。

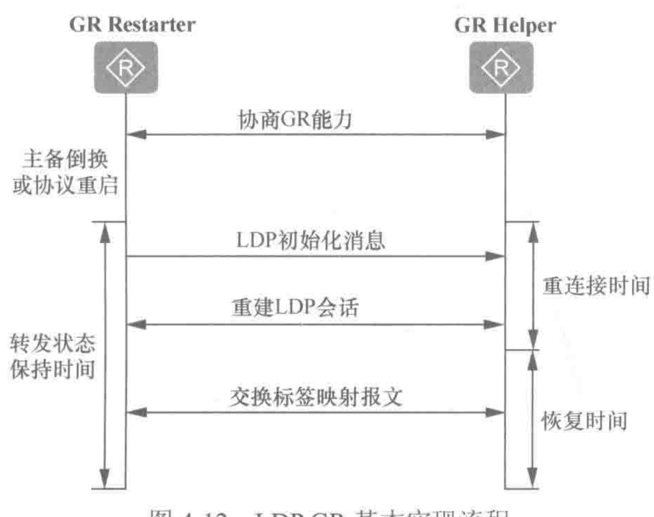


图 4-12　LDP GR 基本实现流程

（5）在恢复定时器超时后，GR Helper 会删除所有标记为 Stale 的 MPLS 转发表项。在 GR Restarter 上的转发状态保持定时器超时后，GR Restarter 删除标记为 Stale 的标签转发表项，结束 GR。

4.5.2　配置 LDP GR

GR Restarter 和 GR Helper 的 LDP GR 功能配置方法是一样的，最主要是只是使能 LDP GR 功能，另外可选配置 4.5.1 节所介绍的 LDP GR 涉及的三个定时器参数（一般可直接采用缺省配置即可），具体配置步骤见表 4-14。但 LDP GR 功能的实现要借助于 IGP GR 功能，所以在配置 LDP GR 之前，需完成以下任务。

- 配置本地 LDP 会话。
- 配置 IGP GR 功能。

IGP GR 方面在此仅介绍在 MPLS 骨干网中最常应用的 OSPF 和 IS-IS 协议的 GR 功能配置。OSPF GR 的配置步骤见表 4-15，IS-IS GR 的配置步骤见表 4-16。

配置好后，可执行 **display mpls graceful-restart** 命令查看 MPLS 相关所有协议的 GR 信息。执行 **display mpls ldp event gr-helper** 命令查看 GR Helper 的相关信息。

表 4-14　LDP GR 的配置步骤

步骤	命令	说明
1	**system-view** 例如：<Huawei> **system-view**	进入系统视图
2	**mpls ldp** 例如：[Huawei] **mpls ldp**	进入 MPLS-LDP 视图

（续表）

步骤	命令	说明
3	**graceful-restart** 例如：[Huawei-mpls-ldp] **graceful-restart**	使能 LDP GR 功能。需要在 GR Restarter 及其邻居节点分别使能。使能或禁止 GR 功能都会导致所有 LDP 实例的会话重建。 缺省情况下，LDP GR 功能未使能，可用 **undo graceful-restart** 命令去使能 LDP GR 功能
4	**graceful-restart timer reconnect** *time* 例如：[Huawei-mpls-ldp] **graceful-restart timer reconnect** 200	（可选）配置 LDP 会话重连接定时器的值，整数形式，取值范围是 3～3600，单位是 s。**仅 AR3260-S 设备支持该命令，也仅需要在 GR Restarter 上配置。** GR Restarter 发生主备倒换后，GR Helper 检测到和 GR Restarter 的 LDP 会话失败，将启动 Reconnect 定时器，等待 LDP 会话的重新建立。 • 如果 Reconnect 定时器超时，GR Helper 和 GR Restarter 之间的 LDP 会话还没有建立，则 GR Helper 立即删除与 GR Restarter 相关的 MPLS 转发表项，退出 GR Helper 流程。 • 如果 Reconnect 定时器超时前，LDP 会话重新建立完成，则 GR Helper 删除该定时器，同时启动 Recovery 定时器。 LDP GR 协商 LDP 会话重连时间时，会取本地配置的邻居存活时间的值和邻居发送的重连接定时器的值的较小值，作为本地实际生效的重连接定时器的值。 缺省情况下，LDP 会话重连接定时器的值为 300s，可用 **undo graceful-restart timer reconnect** 命令恢复缺省设置
5	**graceful-restart timer recovery** *time* 例如：[Huawei-mpls-ldp] **graceful-restart timer recovery** 330	（可选）配置 LSP 恢复定时器的值，整数形式，取值范围是 3～3600，单位是 s。仅需在 GR Helper 上配置 LDP 会话重新建立后，GR Helper 启动 Recovery 定时器，等待 LSP 的恢复。 • 如果 Recovery 定时器超时，GR Helper 认为邻居 GR 结束，未恢复的 LSP 被删除。 • 如果 Recovery 定时器超时之前，所有 LSP 已经恢复，则也要等到该定时器超时后 GR Helper 才认为邻居 GR 结束。 在存在大量路由的网络中，网络故障时，为了防止缺省 300s 内无法恢复所有 LSP，可以配置此命令，调大 LSP 恢复定时器的值。LDP GR 协商 LSP 恢复时间时，会取本地配置的 LSP 恢复定时器的值和邻居发送的 LSP 恢复定时器的值的较小值，作为本地实际生效的 LSP 恢复定时器的值。 缺省情况下，LSP 恢复定时器的值为 300s，可用 **undo graceful-restart timer recovery** 命令恢复缺省配置
6	**graceful-restart timer neighbor-liveness** *time* 例如：[Huawei-mpls-ldp] **graceful-restart timer neighbor-liveness** 500	（可选）配置邻居存活定时器的值，整数形式，取值范围是 3～3600，单位是 s。仅需在 GR Helper 上配置。 LDP GR 协商 LDP 会话重连时间时，会取 GR Helper 配置的 Neighbor-liveness 定时器的值和 GR Restarter 配置的 Reconnect 定时器的值中的较小值。一般情况下，建议使用缺省配置。当网络中 LSP 的数量较少时，可以配置较小的邻居存活定时器的值，短时间内结束 GR。 缺省情况下，邻居存活定时器的值为 600s，可用 **undo graceful-restart timer neighbor-liveness** 命令恢复缺省设置

表 4-15 **OSPF GR 的配置步骤**

步骤	命令	说明
1	**system-view** 例如：<Huawei> **system-view**	进入系统视图
2	**ospf** [*process-id*] 例如：[Huawei] **ospf**	进入 OSPF 视图
3	**opaque-capability enable** 例如：[Huawei-ospf-1] **opaque-capability enable**	使能 opaque-LSA 特性，从而 OSPF 进程可以生成 Opaque LSA，并能从邻居设备接收 Opaque LSA。因为 OSPF 中通过 Type-9 类 LSA 对 OSPF GR 支持，所以需要首先使能 OSPF 的 opauqe-LSA 特性。 缺省情况下，禁止 opaque-lsa 能力，可用 **undo opaque-capability** 命令禁止对 Opaque LSA 进行操作。 缺省情况下，禁止 opaque-lsa 能力，可用 **undo opaque-capability** 命令禁止对 Opaque LSA 进行操作
4	**graceful-restart** 例如：[Huawei-ospf-1] **graceful-restart**	使能 OSPF GR 特性。使能 OSPF GR 功能重启后，Restarter 路由器和 Helper 路由器之间重新建立邻居关系，交换路由信息并同步数据库，更新路由表和转发表，从而实现 OSPF 快速收敛，保持流量不中断，维护网络拓扑稳定。 缺省情况下，关闭 OSPF GR 功能，可用 undo graceful-restart 命令关闭 OSPF GR 功能。

表 4-16 **IS-IS GR 的配置步骤**

步骤	命令	说明
1	**system-view** 例如：<Huawei> **system-view**	进入系统视图
2	**isis** [*process-id*] 例如：[Huawei] **isis**	进入 IS-IS 视图
3	**graceful-restart** 例如：[Huawei-isis-1] **graceful-restart**	使能 IS-IS 协议的 GR 能力。通过在设备运行本命令，可以使能 IS-IS 进程的平滑重启功能，该设备将其重启状态通知给邻居，允许邻居维持邻接关系而保持流量转发不中断。 缺省情况下，未使能 IS-IS 协议的 GR 能力，可用 **undo graceful-restart** 命令去使能 IS-IS 进程的 GR 功能
4	**graceful-restart no-impact-holdtime** 例如：[Huawei-isis-1] **graceful-restart no-impact-holdtime**	使 IS-IS 邻居的 Holdtime 不受 GR 影响，保持原来的数据。 在 IS-IS 网络中，如果一端配置了 GR，则邻居会自动刷新邻居保持时间 holdtime，如果原 holdtime 值小于 60s 则刷新为 60s，否则保留原 holdtime 值，即刷新后邻居的 holdtime 值最小为 60s。因此，在非 GR 期间，链路一端发生了故障，则另一端至少需要 60s 才能感知故障，在此期间可能会造成大量的丢包，降低了网络的安全性和可靠性。基于这种情况，可以通过配置本命令解决该问题，在配置 IS-IS GR 后，仍然可以快速检测邻居状态，实现网络快速收敛。 缺省情况下，配置 IS-IS GR 后，邻居的 holdtime 值如果小于 60s，则修改为 60s，否则保持原 holdtime 值，可用 **undo graceful-restart no-impact-holdtime** 命令恢复缺省配置

（续表）

步骤	命令	说明
5	**graceful-restart interval** *interval-value* 例如：[Huawei-isis-1] **graceful-restart interval** 120	（可选）配置 IS-IS GR 过程中 T3 定时器的时间，整数形式，取值范围是 30～1800，单位是 s。 IS-IS GR 根据重启类型的不同，可以分为 Restarting 和 Starting 两类。其中，Restarting 是指主备倒换和重启 IS-IS 进程引起的 GR 过程，Starting 指 IS-IS 路由器重启引起的 GR 过程。 在 Restarting 过程中，Restarter 设备进行协议重启后会同时启动 T1、T2 和 T3 定时器。T1 定时器用来控制 Restarter 设备收到 Helper 设备发送的确认 GR 的 LSP 报文的时间，T2 定时器是系统等待 LSDB 同步的最长时间，T3 定时器控制 GR 完成的时间。正常的 GR 过程中，当 Level-1 和 Level-2 都完成了 LSDB 同步后，取消 T3 定时器。如果 T3 定时器超时仍未完成 LSDB 同步，则 GR 失败。通过本命令配置使 Restarter 设备的邻居将 T3 定时器的时间设置为邻居保持时间，避免 GR 期间邻居断连造成整个网络路由的重新计算，出现 LSDB 未完成同步而 T3 定时器超时导致 GR 失败的情况。 缺省情况下，T3 定时器为 300s，可用 **undo graceful-restart interval** 命令恢复 T3 定时器的缺省值。建议保持该缺省值。
6	**graceful-restart suppress-sa** 例如：[Huawei-isis-1] **graceful-restart suppress-sa**	（可选）配置 GR Restarter 来抑制重启 TLV 的 SA（Suppress-Advertisement，抑制发布）位。 第一次启动（不包括 GR 后）的路由器不会对转发状态进行维护。如果该路由器不是第一次启动，则它前一次运行时生成的 LSP 可能还存在于网络中其他路由器的 LSP 数据库中。但由于路由器启动时 LSP 分片的序列号也被重新初始化，网络中其他路由器保存的 LSP 拷贝可能会比该路由器启动后新产生的 LSP 看上去更“新”。这将导致网络中出现暂时的“黑洞”（black-hole），并一直持续到该路由器重新生成自己的 LSP 且以最高序列号将它们发布出去。如果该路由器的邻居在路由器启动过程中抑制发布邻接关系到此路由器，直到该路由器将更新的 LSP 发布出去，上述情况也可以避免。 缺省情况下，不对 SA 位进行抑制，可用 **undo graceful-restart suppress-sa** 命令恢复为缺省状态

4.5.3 LDP GR 配置示例

如图 4-13 所示，部署了 MPLS LDP 业务，节点 LSRA、LSRB 和 LSRC 都为单主控（正常工作时只有一个主控板处于工作状态）设备。在主备倒换过程或系统升级中，邻居会因为会话进入 Down 状态而删除与本设备相关的 LSP 而导致业务、流量短时间中断。现要求在主备倒换过程或系统升级中，邻居不会因为会话进入 Down 状态而删除 LSP，以实现流量短时间不中断。

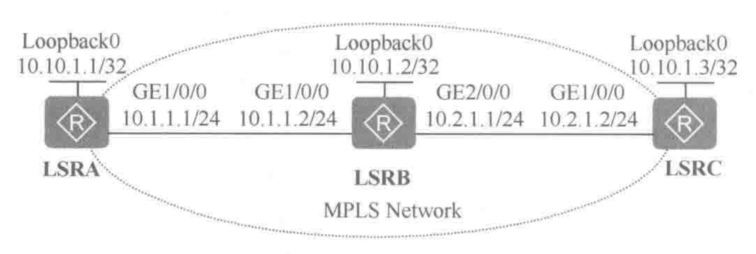


图 4-13 LDP GR 配置示例的拓扑结构

1. 基本配置思路分析

很明显本示例仅通过配置 LDP GR 即可实现此需求。但在配置 LDP GR 之前也需要配置好骨干网各节点的 MPLS 基本能力和 LDP 本地会话，以及使能对应的 IGP GR 功能。据此可得出本示例如下的基本配置思路。

（1）配置各设备的接口 IP 地址和 OSPF 路由，实现骨干网三层互通。

（2）配置各设备 MPLS 基本能力和 LDP 本地会话。

（3）使能各设备的 OSPF GR 功能。

（4）使能各设备的 LDP FRR 功能。

2. 具体配置步骤

（1）配置各设备接口的 IP 地址和 OSPF 路由。

LSRA 上的配置。

```
<Huawei> system-view
[Huawei] sysname LSRA
[LSRA] interface loopback 0
[LSRA-LoopBack0] ip address 10.10.1.1 32
[LSRA-LoopBack0] quit
[LSRA] interface gigabitethernet 1/0/0
[LSRA-GigabitEthernet1/0/0] ip address 10.1.1.1 24
[LSRA-GigabitEthernet1/0/0] quit
[LSRA] ospf 1
[LSRA-ospf-1] area 0
[LSRA-ospf-1-area-0.0.0.0] network 10.10.1.1 0.0.0.0
[LSRA-ospf-1-area-0.0.0.0] network 10.1.1.0 0.0.0.255
[LSRA-ospf-1-area-0.0.0.0] quit
[LSRA-ospf-1] quit
```

LSRB 上的配置。

```
<Huawei> system-view
[Huawei] sysname LSRB
[LSRB] interface loopback 0
[LSRB-LoopBack0] ip address 10.10.1.2 32
[LSRB-LoopBack0] quit
[LSRB] interface gigabitethernet 1/0/0
[LSRB-GigabitEthernet1/0/0] ip address 10.1.1.2 24
[LSRB-GigabitEthernet1/0/0] quit
[LSRB] interface gigabitethernet 2/0/0
[LSRB-GigabitEthernet2/0/0] ip address 10.2.1.2 24
[LSRB-GigabitEthernet2/0/0] quit
[LSRB] ospf 1
[LSRB-ospf-1] area 0
[LSRB-ospf-1-area-0.0.0.0] network 10.10.1.2 0.0.0.0
[LSRB-ospf-1-area-0.0.0.0] network 10.1.1.0 0.0.0.255
[LSRB-ospf-1-area-0.0.0.0] network 10.2.1.0 0.0.0.255
[LSRB-ospf-1-area-0.0.0.0] quit
[LSRB-ospf-1] quit
```

LSRC 上的配置。

```
<Huawei> system-view
[Huawei] sysname LSRC
[LSRC] interface loopback 0
[LSRC-LoopBack0] ip address 10.10.1.3 32
```

```
[LSRC-LoopBack0] quit
[LSRC] interface gigabitethernet 1/0/0
[LSRC-GigabitEthernet1/0/0] ip address 10.2.1.2 24
[LSRC-GigabitEthernet1/0/0] quit
[LSRC] ospf 1
[LSRC-ospf-1] area 0
[LSRC-ospf-1-area-0.0.0.0] network 10.10.1.3 0.0.0.0
[LSRC-ospf-1-area-0.0.0.0] network 10.2.1.0 0.0.0.255
[LSRC-ospf-1-area-0.0.0.0] quit
[LSRC-ospf-1] quit
```

以上配置完成后，在各节点上执行 **display ip routing-table** 命令，可以看到相互之间都学到了彼此的路由。

（2）配置各设备的 MPLS 基本能力和 LDP 本地会话。

LSRA 上的配置。

```
[LSRA] mpls lsr-id 10.10.1.1
[LSRA] mpls
[LSRA-mpls] quit
[LSRA] mpls ldp
[LSRA-mpls-ldp] quit
[LSRA] interface gigabitethernet 1/0/0
[LSRA-GigabitEthernet1/0/0] mpls
[LSRA-GigabitEthernet1/0/0] mpls ldp
[LSRA-GigabitEthernet1/0/0] quit
```

LSRB 上的配置。

```
[LSRB] mpls lsr-id 10.10.1.2
[LSRB] mpls
[LSRB-mpls] quit
[LSRB] mpls ldp
[LSRB-mpls-ldp] quit
[LSRB] interface gigabitethernet 1/0/0
[LSRB-GigabitEthernet1/0/0] mpls
[LSRB-GigabitEthernet1/0/0] mpls ldp
[LSRB-GigabitEthernet1/0/0] quit
[LSRB] interface gigabitethernet 2/0/0
[LSRB-GigabitEthernet2/0/0] mpls
[LSRB-GigabitEthernet2/0/0] mpls ldp
[LSRB-GigabitEthernet2/0/0] quit
```

LSRC 上的配置。

```
[LSRC] mpls lsr-id 10.10.1.3
[LSRC] mpls
[LSRC-mpls] quit
[LSRC] mpls ldp
[LSRC-mpls-ldp] quit
[LSRC] interface gigabitethernet 1/0/0
[LSRC-GigabitEthernet1/0/0] mpls
[LSRC-GigabitEthernet1/0/0] mpls ldp
[LSRC-GigabitEthernet1/0/0] quit
```

（3）使能各设备的 OSPF GR 功能。

LSRA 上的配置。

```
[LSRA]ospf
[LSRA-ospf-1] opaque-capability enable  #---使能 opaque-LSA 特性，从而 OSPF 进程可以生成 Opaque LSA，并能从邻居设备接收 Opaque LSA
```

```
[LSRA-ospf-1] graceful-restart   #---使能 OSPF GR 特性
[LSRA-ospf-1] quit
```

LSRB 上的配置。

```
[LSRB]ospf
[LSRB-ospf-1] opaque-capability enable
[LSRB-ospf-1] graceful-restart
[LSRB-ospf-1] quit
```

LSRC 上的配置。

```
[LSRC]ospf
[LSRC-ospf-1] opaque-capability enable
[LSRC-ospf-1] graceful-restart
[LSRC-ospf-1] quit
```

（4）使能各设备的 LDP GR 功能。各定时器均采用缺省值。

#　LSRA 上的配置。

```
[LSRA] mpls ldp
[LSRA-mpls-ldp] graceful-restart
Warning: All the related sessions will be deleted if the operation is performed
!Continue? (y/n)y
[LSRA-mpls-ldp] quit
```

#　LSRB 上的配置。

```
[LSRB] mpls ldp
[LSRB-mpls-ldp] graceful-restart   #---使能 LDP GR 特性
Warning: All the related sessions will be deleted if the operation is performed
!Continue? (y/n)y
[LSRB-mpls-ldp] quit
```

#　LSRC 上的配置。

```
[LSRC] mpls ldp
[LSRC-mpls-ldp] graceful-restart
Warning: All the related sessions will be deleted if the operation is performed
!Continue? (y/n)y
[LSRC-mpls-ldp] quit
```

3. 配置结果验证

完成上述配置后，在 LSR 上执行 **display mpls ldp session verbose** 命令，可以看到“**Session FT Flag**”字段的值为“**On**”，表示设备支持 LDP GR。以下是在 LSRA 上执行该命令的输出示例。

```
[LSRA]display mpls ldp session verbose

                    LDP Session(s) in Public Network
 ------------------------------------------------------------------------
 Peer LDP ID         : 10.10.1.2:0              Local LDP ID    : 10.10.1.1:0
 TCP Connection      : 10.10.1.1 <- 10.10.1.2
 Session State       : Operational              Session Role    : Passive
 Session FT Flag : On                  MD5 Flag          : Off
 Reconnect Timer : 0 Sec                        Recovery Timer : 300 Sec
 Keychain Name       : ---

 Negotiated Keepalive Hold Timer       : 45 Sec
 Configured Keepalive Send Timer       : ---
 Keepalive Message Sent/Rcvd           : 1/1 (Message Count)
 Label Advertisement Mode              : Downstream Unsolicited
 Label Resource Status(Peer/Local) : Available/Available
```

```
Session Age                              : 0000:00:00  (DDDD:HH:MM)
Session Deletion Status                  : No

Capability:
  Capability-Announcement                : Off
 mLDP P2MP Capability                    : Off
 mLDP MBB Capability                     : Off

Outbound&Inbound Policies applied : NULL
Addresses received from peer: (Count: 3)
10.10.1.2              10.1.1.2              10.2.1.1

------------------------------------------------------------------------------
```

或者在 LSR 上执行 **display mpls ldp peer verbose** 命令，可以看到“**Peer FT Flag**”字段的值为“**On**”，表示设备支持 LDP GR。以下是在 LSRA 上执行该命令的输出示例。

```
[LSRA] display mpls ldp peer verbose

 LDP Peer Information in Public network
 ------------------------------------------------------------------------------
 Peer LDP ID            : 10.10.1.2:0
 Peer Max PDU Length : 4096              Peer Transport Address : 10.10.1.2
 Peer Loop Detection : Off               Peer Path Vector Limit : ----
 Peer FT Flag           : On             Peer Keepalive Timer    : 45 Sec
 Recovery Timer         : 300 Sec        Reconnect Timer          : 0 Sec
 Peer Type              : Local

 Peer Label Advertisement Mode : Downstream Unsolicited
 Peer Discovery Source         : GigabitEthernet1/0/0
 Peer Deletion Status          : No
 Capability-Announcement       : Off
 Peer mLDP P2MP Capability     : Off
 Peer mLDP MBB Capability      : Off
 ------------------------------------------------------------------------------
```

4.6 LDP 安全机制配置与管理

为了提高邻居间建立 LDP 会话的安全性，防止非法设备接入骨干网，防止非法 LDP 报文攻击，MPLS 提供了三种保护机制：LDP MD5 认证、LDP Keychain 认证和 LDP GTSM。

LDP Keychain 认证是比 LDP MD5 认证更安全的加密认证，对于同一邻居，只能选择其中一个加密认证；而 LDP GTSM 用来防止设备受到非法 LDP 报文的攻击，其可以与前面两种 LDP 邻居认证配合使用。

4.6.1 LDP 安全机制简介

前面说到，MPLS 提供了三种保护机制：LDP MD5 认证、LDP Keychain 认证和 LDP GTSM，下面分别予以介绍。

1. LDP MD5 认证

MD5 是 RFC1321 定义的国际标准摘要密码算法，其典型应用是针对一段信息计算

出对应的信息摘要，从而防止信息被篡改，这方面的原理已在《华为 VPN 学习指南》一书中有详细介绍。

MD5 信息摘要是通过不可逆的字符串变换算法产生的，结果唯一。因此，不管信息内容在传输过程中发生任何形式的改变，只要重新计算就会产生不同的信息摘要（现在也已发现有碰撞的现象，但概率极小），接收端就可以由此判定收到的是一个不正确的报文。

LDP MD5 应用其对同一信息段产生唯一摘要信息的特点来实现 LDP 报文防篡改校验，比一般意义上 TCP 校验和更为严格。它在 LDP 报文传输中的实现过程如下。

（1）LDP 会话消息在经 TCP 协议发出前，会在 TCP 头后面填充一个经 MD5 算法计算后的信息摘要。而这个信息摘要就是把 TCP 头部、LDP 会话消息以及用户设置的密码一起作为原始信息，通过 MD5 算法计算出。

（2）当接收端收到这个 TCP 报文时，首先会取得报文的 TCP 头部、信息摘要、LDP 会话消息，并结合 TCP 头部、LDP 会话消息以及本地保存的密码，再利用 MD5 计算出信息摘要，然后与报文携带的信息摘要进行比较，从而检验报文是否被篡改过。

在用户设置密码时有明文和密文两种形式选择，这里的明文、密文是对用户设置的密码在配置文件中的记录形式。明文就是直接在配置文件中记录用户设置的字符串，密文就是在配置文件中记录经过特殊算法加密后的字符串。但无论用户选择密码记录形态是明文还是密文形式，**参与摘要计算时都是直接使用用户输入的字符串**，也就是说加密算法计算出的密码并不会参与 MD5 摘要计算。这样一来，即使由于各厂商所采用的明文、密文的转化算法不兼容，也能进行唯一的 MD5 运算，使各厂商的明文、密文转换算法在 MD5 消息摘要计算中相互透明（就是与明文、密文转换算法无关）。

2. LDP Keychain 认证

Keychain 是一种增强型加密算法，类似于 MD5，Keychain 也是针对同一段信息计算出对应的信息摘要，实现 LDP 报文防篡改校验。

Keychain 允许用户定义一组密码，形成一个密码串，并且分别为每个密码指定加解密算法（包括 MD5、SHA-1 等）及密码使用的有效时间。在收发报文时，系统会按照用户的配置选出一个当前有效的密码，并按照与此密码相匹配的加密解密算法以及密码的有效时间，进行发送时加密和接收时解密报文。也就是在不同时间，所使用的密码是不同的，所使用的加/解密算法也可能不同，这样比固定使用相同的密码、相同的加/解密算法具有更高的安全性。此外，系统可以依据密码使用的有效时间，自动完成有效密码的切换，避免了长时间不更改密码导致的密码易破解问题。

Keychain 的密码、所使用的加/解密算法，以及密码使用的有效时间可以单独配置，形成一个 Keychain 配置节点，每个 Keychain 配置节点至少需要配置一个密码，并指定加解密算法。

3. LDP GTSM

GTSM（Generalized TTL Security Mechanism，通用 TTL 安全保护机制）是一种通过检查 IP 报头中的 TTL 值是否在一个预先定义好的范围内来实现对 IP 业务进行保护的机制。使用 GTSM 有两个前提。

■ 设备之间正常报文的 TTL 值是确定的。

■ 报文的 TTL 值很难被修改，否则就无法根据 TTL 值进行保护。

LDP GTSM 是 GTSM 在 LDP 方面的具体应用。GTSM 通过判定报文的 TTL 值，确定报文是否有效，从而保护设备免受攻击。LDP GTSM 是对相邻或相近（基于只要跳数确定的原则）设备间的 LDP 消息报文应用此种机制。用户预先在各设备上设定好针对其他设备报文的有效范围，使能 GTSM，这样当相应设备之间应用 LDP 时，如果 LDP 消息报文的 TTL 不符合之前设置的范围要求，设备就认为此报文为非法攻击报文予以丢弃，进而实现对上层协议的保护。

4.6.2 配置 LDP MD5 认证

为了提高 LDP 会话连接的安全性，可以对 LDP 使用的 TCP 连接配置 MD5 认证。LDP 会话的两个对等体可以配置不同的加密方式，但是密码必须相同。

MD5 算法配置简单，具体见表 4-17。配置后生成单一密码，需要人为干预才可以切换密码，适用于需要短时间加密的网络。

注意 对于同一邻居，在配置 Keychain 认证后，不能再配置 MD5 认证；同样，在配置 MD5 认证后，不能再配置 Keychain 认证。但 Keychain 认证比 MD5 认证方式更安全。

另外，配置 LDP MD5 认证可能会导致 LDP 会话重建，与原来会话相关的 LSP 将被删除，造成 MPLS 业务中断。

表 4-17 **LDP MD5 认证的配置步骤**

步骤	命令	说明
1	**system-view** 例如：<Huawei> **system-view**	进入系统视图
2	**mpls ldp** 例如：[Huawei] **mpls ldp**	进入 MPLS-LDP 视图
3	**md5-password** { **plain** | **cipher** } *peer-lsr-id password* 例如：[Huawei-mpls-ldp] **md5-password cipher** 2.2.2.2 Huawei-123	使能 MD5 认证，并配置认证密码。命令中的参数和选项说明如下。 • **Plain**：二选一选项，以明文形式显示配置的密码。此时，密码将以明文形式保存在配置文件中，低级别登录的用户可以通过查看配置方式获取密码，造成安全隐患。 • **cipher**：二选一选项，以密文形式显示配置的密码。建议使用 **cipher** 选项，将密码加密保存。 • *peer-lsr-id*：对等体的 LSR ID，用于标识要采用与指定对等体采用 MD5 认证方式。 • *password*：密码字符串，不支持空格。如果采用明文形式，只能输入密码长度范围是 1～255 的字符串。如果采用密文形式，只能输入密码长度范围是 20～392 的字符串。当输入的字符串两端使用双引号时，可在字符串中输入空格。 改变对等体密码后，LDP 会重新创建会话，与原来会话相关的 LSP 将被删除。 缺省情况下，LDP 对等体之间不进行 MD5 认证，**undo md5-password** [**plain** | **cipher**] *peer-lsr-id* 取消与指定对等体采用 MD5 认证方式

4.6.3 配置 LDP Keychain 认证

为了提高 LDP 会话连接的安全性，可以对 LDP 使用的 TCP 连接配置 Keychain（密钥链）认证。Keychain 具有一组密码，可以根据配置在不同时间自动切换使用不同的密码，使认证的安全性得到极大提高。但是这种认证的配置过程较为复杂，适用于对安全性能要求比较高的网络。配置 LDP Keychain 认证可能会导致 LDP 会话重建，与原来会话相关的 LSP 将被删除，造成 MPLS 业务中断。

LDP Keychain 认证的配置也很简单，复杂之处主要体现在要先配置好 LDP Keychain 认证所调用的密钥链。配置 Keychain 时，Key 代表 Keychain 认证规则，每一个 Key 包括：认证算法、认证加密的密钥、活跃的发送时间、活跃的接收时间。一个 Keychain 中最多可以有 64 个 Key。

说明　同一个 Keychain 中各 Key 必须唯一。不同的 Keychain 中 Key ID 可以相同。同一 Keychain 中不能有多个发送 Key 同时生效，否则应用程序无法选择应用哪个发送 Key 进行加密。但是同一 Keychain 中可以有多个接收 Key 同时生效，接收端会根据接收到的 Key ID，选择 ID 相同的活跃的接收 Key 进行解密。

如果发送端的 Key 发生变更，接收端的 Key 也需要变更。由于网络上时钟可能不同步，在接收端和发送端变更 Key 时，有可能存在时间延迟。在延迟的时间范围内会造成数据丢失。为了实现两端 Key 变更时不丢包，可以配置接收容忍时间。接收容忍时间只对接收端的 Key 有效。接收容忍时间将导致接收的起始和终止的时间延长。

如果在某个时间段管理员没有配置 Key，此时将没有活跃的发送 Key。在该时间段，应用程序将没有认证的交互。为了避免这种情况，可以配置缺省发送 Key。任何存在的 Key 都可以被指定为缺省发送 Key，在一个 Keychain 中只能有一个缺省发送 Key。在没有其他活跃的发送 Key 时，应用程序将使用缺省发送 Key 作为活跃的发送 Key。

Keychain 的配置步骤见表 4-18，LDP Keychain 认证的配置步骤见表 4-19。

表 4-18　**Keychain 的配置步骤**

步骤	命令	说明
1	**system-view** 例如：<Huawei> **system-view**	进入系统视图
2	**keychain** *keychain-name* **mode** { **absolute** | **periodic** { **daily** | **weekly** | **monthly** | **yearly** } } 例如：[Huawei] **keychain huawei mode absolute**	创建 Keychain，并进入 Keychain 视图。命令中的参数和选项说明如下。 • *keychain-name*：指定 Keychain 的名称，字符串形式，长度范围是 1～47，不区分大小写。字符不包括问号和空格，但是当输入的字符串两端使用双引号时，可在字符串中输入空格。LDP 通过 Keychain 名称调用 Keychain。 • **mode**：指定 Keychain 的生效的时间模。 • **absolute**：二选一选项，指定 Keychain 以绝对时间生效，不以周期形式生效。此时，Keychain 仅在一个时间段内一次性生效，超出该时间段 Keychain 永远不再生效。

（续表）

步骤	命令	说明
2	**keychain** *keychain-name* **mode** { **absolute** | **periodic** { **daily** | **weekly** | **monthly** | **yearly** } } 例如：[Huawei] **keychain** huawei **mode absolute**	• **periodic**：二选一选项，指定 Keychain 以周期形式生效。此时，Keychain 仅在一个时间段内周期性生效，超出该时间段后，下一个周期的该时间段继续生效。 • **daily**：多选一选项，指定 Keychain 以日方式周期生效。 • **weekly**：多选一选项，指定 Keychain 以星期方式周期生效。 • **monthly**：多选一选项，指定 Keychain 以月方式周期生效。 • **yearly**：多选一选项，指定 Keychain 以年方式周期生效。 【说明】每个 Keychain 中由多个 Key 组成，每一 Key 需要对应配置一个认证算法，不同的 Key 在不同时间段活跃，从而实现 Keychain 认证算法的动态切换。配置 Key 的发送和接收时间时，需要和 Keychain 的时间模式一致。 缺省情况下，没有配置 Keychain，可用 **undo keychain** *keychain-name* 命令删除 Keychain 配置
3	**key-id** *key-id* 例如：[Huawei-keychain-huawei] **key-id** 1	创建 key-id，并进入 key-id 视图。参数 *key-id* 为 key-id 的值，用来唯一标识 Keychain 中的 Key，整数形式，取值范围是 0～63。 缺省情况下，没有配置 Key-id，可用 **undo key-id** *key-id* 命令删除指定的 Key-id
4	**algorithm** { **hmac-md5** | **hmac-sha-256** | **hmac-sha1-12** | **hmac-sha1-20** | **md5** | **sha-1** | **sha-256** | **simple** } 例如：[Huawei-keychain-huawei-keyid-1] **algorithm sha-256**	配置 key 采用的认证加密算法。命令中的选项说明如下。 • **hmac-md5**：多选一选项，指定采用 HMAC-MD5 认证算法对报文进行加密和认证，产生 128bit 的信息摘要，安全性较低。 • **hmac-sha-256**：多选一选项，指定采用 HMAC-SHA-256 认证算法对报文进行认证，产生 128bit 的信息摘要。 • **hmac-sha1-12**：多选一选项，指定采用 HMAC-SHA1-12 认证算法对报文进行加密和认证，产生 160bit 的摘要信息。 • **hmac-sha1-20**：多选一选项，指定采用 HMAC-SHA1-20 认证算法对报文进行加密和认证，产生 160bit 的摘要信息。 • **md5**：多选一选项，指定采用 MD5 认证算法对报文进行加密和认证，产生 128bit 的消息摘要，安全性较低。 • **sha-1**：多选一选项，指定采用 SHA-1 认证算法对报文进行加密和认证，产生 160bit 的消息摘要，安全性较低。 • **sha-256**：多选一选项，指定采用 SHA-256 认证算法对报文进行认证。 • **simple**：多选一选项，指定采用配置的密钥对报文进行认证，不安全，不建议采用。 【注意】发送方 Key 的认证加密算法必须和接收方 Key 的认证加密算法一致，否则将导致应用协议因认证不通过而断开连接。不配置认证算法，Key 将处于非活跃状态。 缺省情况下，没有配置认证算法，可用 undo algorithm 命令用来删除 Key 的认证算法
5	**key-string** { **plain** *plain-text* | [**cipher**] *cipher-text* } 例如：[Huawei-keychain-huawei-keyid-1] **key-string cipher** Huawei@1234	配置 Key 的认证加密的密钥。命令中的参数说明如下。 • **plain** *plain-text*：二选一参数，指定明文密钥，字符串形式，区分大小写，可以是字母或数字，长度范围是 1～255 字节。当密码包含空格时，密码需要加双引号，并且只能有这一个双引号。以明文方式输入，明文方式显示。

（续表）

步骤	命令	说明
5	**key-string** { **plain** *plain-text* | [**cipher**] *cipher-text* } 例如：[Huawei-keychain-huawei-keyid-1] **key-string cipher** Huawei@1234	• **cipher**：可选项，指定采用密文口令类型。需要用户保存好明文形式的密钥，方便遗忘时找回密钥。 • *cipher-text*：二选一参数，指定密文密钥，字符串形式，区分大小写，可以是字母或数字。可以支持明文或密文形式输入，以密文形式显示。可以输入 1～255 的明文字符串，也可以输入 20～392 的密文字符串。当密码包含空格时，密码需要加双引号，并且只能有这一个双引号。 缺省情况下，没有配置认证的密钥，可用 **undo key-string** 命令删除 Keychain 认证的密钥
	以下是根据步骤 2 中选择的不同时间模式配置发送时间	
6	**send-time** *start-time start-date* { **duration** { *duration-value* | **infinite** } | **to** *end-time end-date* } 例如：[Huawei-keychain-huawei-keyid-1] **send-time** 14:52 2017-10-1 **to** 14:52 2040-10-1	（多选一）当选择 **absolute** 时间模式时，配置 Key 发送报文生效的时间段。命令中的参数说明如下。 • *start-time*：指定发送的开始时间，HH:MM 方式，取值范围是 00:00～23:59。 • *duration-value*：二选一参数，指定发送报文的持续时间，取值范围是 1～26280000，单位为 min。 • **infinite**：二选一选项，指定 Keychain 发送报文从配置的开始时间起永久活跃。 • **to** *end-time end-date*：二选一参数，指定发送的结束时间，HH:MM 方式，取值范围是 00:00～23:59。结束时间必须大于开始时间。 缺省情况下，没有配置 Key 的发送时间段，可用 **undo send-time** 命令删除 Key 的发送报文生效时间段
	send-time daily *start-time* **to** *end-time* 例如：[Huawei-keychain-huawei-keyid-1] **send-time daily** 14:52 **to** 18:10	（多选一）当选择 **periodic daily** 时间模式时，配置 Key 发送报文生效的时间段。参数说明参见上面 **absolute** 时间模式中 send-time 中对应的参数说明。 缺省情况下，有配置 Key 的发送时间段，可用 **undo send-time** 命令删除 key 的发送报文生效时间段
	send-time day { *start-day-name* **to** *end-day-name* | *day-name &<1-7>* } 例如：[Huawei-keychain-huawei-keyid-1] **send-time day mon to fri**	（多选一）当选择 periodic weekly 时间模式时，配置 Key 发送报文生效的时间段。命令中的参数说明如下。 • *start-day-name* **to** *end-day-name*：二选一参数，指定 Keychain 每周发送报文生效的起始和结束日期，取值范围是：mon（星期一）、tue（星期二）、wed（星期三）、thu（星期四）、fri（星期五）、sat（星期六）和 sun（星期日）。 • *day-name &<1-7>*：二选一参数，指定 Keychain 每周发送报文生效的日期，取值范围是：mon（星期一）、tue（星期二）、wed（星期三）、thu（星期四）、fri（星期五）、sat（星期六）和 sun（星期日），可以取其中任意一个或多个日期。 缺省情况下，有配置 Key 的发送时间段，可用 **undo send-time** 命令删除 Key 的发送报文生效时间段

（续表）

步骤	命令	说明
6	**send-time date** { *start-date-value* **to** *end-date-value* | *date-value* &<1-31> } 例如：[Huawei-keychain-huawei-keyid-1] **send-time date** 1 **to** 30	（多选一）当选择 **periodic monthly** 时间模式时，配置 Key 发送报文生效的时间段。命令中的参数说明如下。 ● *start-date-value* **to** *end-date-value*：二选一参数，指定 Keychain 每月发送报文生效的起始和结束日期，起始日期的取值范围是 1～31，结束日期的取值范围是 2～31。结束日期必须大于开始日期。 ● *date-value* &<1-31>：二选一参数，指定 Keychain 每月发送报文生效的日期，取值范围是 1～31，可以取其中任意一个或多个日期。 缺省情况下，有配置 Key 的发送时间段，可用 **undo send-time** 命令删除 Key 的发送报文生效时间段
	send-time month { *start-month-name* **to** *end-month-name* | *month-name* &<1-12> } 例如：[Huawei-keychain-huawei-keyid-1] **send-time month jan to jun**	（多选一）当选择 **periodic yearly** 时间模式时，配置 key 发送报文生效的时间段。命令中的参数说明如下。 ● *start-month-name* **to** *end-month-name*：二选一参数，指定 Keychain 每年发送报文生效的起始和结束月份。起始月份的取值范围是：jan（一月）、feb（二月）、mar（三月）、apr（四月）、may（五月）、jun（六月）、jul（七月）、aug（八月）、sep（九月）、oct（十月）、nov（十一月）和 dec（十二月），结束月份的取值范围为 feb（二月）、mar（三月）、apr（四月）、may（五月）、jun（六月）、jul（七月）、aug（八月）、sep（九月）、oct（十月）、nov（十一月）和 dec（十二月），结束月份必须大于开始月份。 ● *month-name* &<1-12>：二选一参数，指定 Keychain 每年发送报文生效的月份，取值范围是：jan（一月）、feb（二月）、mar（三月）、apr（四月）、may（五月）、jun（六月）、jul（七月）、aug（八月）、sep（九月）、oct（十月）、nov（十一月）和 dec（十二月），可以取其中任意一个或多个月份。 缺省情况下，有配置 Key 的发送时间段，可用 **undo send-time** 命令删除 Key 的发送报文生效时间段
以下是根据步骤 2 中选择的不同时间模式配置接收时间		
7	**receive-time** *start-time start-date* { **duration** { *duration-value* | **infinite** } | **to** *end-time end-date* } 例如：[Huawei-keychain-huawei-keyid-1] **receive-time** 14:52 2017-10-1 **duration infinite**	（多选一）当选择 **absolute** 时间模式时，配置 Key 接收报文生效的时间段。命令中的参数说明参见本表前面 **absolute** 时间模式时 **send-time** 命令中的对应参数说明。 缺省情况下，没有配置 key 的接收时间段，可用 **undo receive-time** 命令删除 Key 的接收报文时间段
	receive-time daily *start-time* **to** *end-time* 例如：[Huawei-keychain-huawei-keyid-1] **receive-time daily** 14:52 **to** 18:10	（多选一）当选择 **periodic daily** 时间模式时，配置 Key 接收报文生效的时间段。命令中的参数说明参见本表前面 **periodic daily** 时间模式时 **send-time** 命令中的对应参数说明。 缺省情况下，没有配置 Key 的接收时间段，可用 **undo receive-time** 命令删除 key 的接收报文时间段

（续表）

<table>
<tr><th>步骤</th><th>命令</th><th>说明</th></tr>
<tr><td rowspan="3">7</td><td>receive-time day { start-day-name to end-day-name | day-name &<1-7> }
例如：[Huawei-keychain-huawei-keyid-1] receive-time day tue to fri</td><td>（多选一）当选择 periodic weekly 时间模式时，配置 key 接收报文生效的时间段。命令中的参数说明参见本表前面 periodic weekly 时间模式时 send-time 命令中的对应参数说明。
缺省情况下，没有配置 Key 的接收时间段，可用 undo receive-time 命令删除 Key 的接收报文时间段</td></tr>
<tr><td>receive-time date { start-date-value to end-date-value | date-value &<1-31> }
例如：[Huawei-keychain-huawei-keyid-1] receive-time date 1、10、12</td><td>（多选一）当选择 periodic monthly 时间模式时，配置 Key 接收报文生效的时间段。命令中的参数说明参见本表前面 periodic monthly 时间模式时 send-time 命令中的对应参数说明。
缺省情况下，没有配置 Key 的接收时间段，可用 undo receive-time 命令删除 key 的接收报文时间段</td></tr>
<tr><td>receive-time month { start-month-name to end-month-name | month-name &<1-12> }
例如：[Huawei-keychain-huawei-keyid-1] receive-time month jan to dec</td><td>（多选一）当选择 periodic yearly 时间模式时，配置 key 接收报文生效的时间段。命令中的参数说明参见本表前面 periodic yearly 时间模式时 send-time 命令中的对应参数说明。
缺省情况下，没有配置 Key 的接收时间段，可用 undo receive-time 命令删除 Key 的接收报文时间段</td></tr>
<tr><td>8</td><td>default send-key-id
例如：[Huawei-keychain-huawei-keyid-1] default send-key-id</td><td>配置该 Key 为缺省发送 Key。
当 Keychain 中不存在发送 Key，或者某个时间段没有活跃的发送 key 时，Keychain 将不能对协议报文进行认证和加密处理，导致应用协议因认证不通过而断开连接。配置缺省的发送 Key 可以保证在没有活跃的 Key 时，Keychain 采用该 Key 对协议报文进行认证和加密，从而保证协议报文的正常通信。
【注意】一个 Keychain 中只能存在一个缺省的发送 Key。
● 当指定的缺省发送 Key 是一个已经存在的 Key 时，缺省发送 Key 直接继承该 Key 的认证加密算法和密钥。
● 当指定的缺省发送 Key 是一个新创建的 Key 时，需要同时配置 Key 的认证加密算法和密钥。
缺省情况下，没有配置缺省发送 Key，可用 undo default send-key-id 命令删除 Keychain 配置的缺省发送 Key</td></tr>
</table>

表 4-19　　**LDP Keychain 认证的配置步骤**

步骤	命令	说明
1	**system-view** 例如：<Huawei> **system-view**	进入系统视图
2	**mpls ldp** 例如：[Huawei] **mpls ldp**	进入 MPLS-LDP 视图
3	**authentication key-chain peer** *peer-id* **name** *keychain-name* 例如：[Huawei-mpls-ldp] **authentication key-chain peer** 2.2.2.2 **name** Huawei	使能 LDP Keychain 认证，并引用配置的 Keychain 名称。命令中的参数说明如下。 ● *peer-id*：指定使用 LDP Keychain 认证建立 LDP 会话的对等体的 ID。 ● *keychain-name*：指定在表 4-18 中所配置的 Keychain 的名称。 缺省情况下，LDP 对等体之间不进行 LDP Keychain 认证，可用 **undo authentication key-chain peer** *peer-id* 命令去使能与指定对等体使用 LDP Keychain 认证建立 LDP 会话

4.6.4　配置 LDP GTSM

GTSM（Generalized TTL Security Mechanism，通用 TTL 安全保护机制）通过判定报文的 TTL 值来确定报文是否有效，从而保护设备免受攻击。在 LDP 对等体上配置 GTSM 功能，通过配置的 TTL 有效范围，对 LDP 对等体间的 LDP 消息报文进行 TTL 检测。如果 LDP 消息报文的 TTL 不符合配置的范围要求，就认为此报文为非法攻击报文予以丢弃，以免 LDP 协议遭到大量伪装报文的攻击，进而实现对上层协议的保护。

GTSM 可以与 LDP MD5 认证或 LDP Keychain 认证结合使用，更加安全。配置 GTSM 的步骤见表 4-20。

表 4-20　　GTSM 的配置步骤

步骤	命令	说明
1	**system-view** 例如：<Huawei> **system-view**	进入系统视图
2	**mpls ldp**	进入 MPLS-LDP 视图
3	**gtsm peer** *ip-address* **valid-ttl-hops** *hops* 例如：[Huawei-mpls-ldp] **gtsm peer** 2.2.2.9 **valid-ttl-hops** 2	配置 LDP GTSM 功能。命令中的参数说明如下。 • **peer** *ip-address*：指定 LDP 对等体的传输地址，即对等体的 LSR-ID • **valid-ttl-hops** *hops*：允许指定对等体离本地设备的最大有效跳数，整数形式，取值范围是 1～255。如果将 *hops* 设置为 GTSM 功能允许的最大有效跳数，当 LDP 对等体发来报文的 TTL 值在[255–hops+1,255]范围内，则接收该报文，否则丢弃该报文。 LDP 报文初始的 TTL 值通常为 255（也可从 IP 报头中的 TTL 字段复制得到），每经过一跳减 1。 缺省情况下，没有在任何 LDP 对等体上配置 GTSM 功能，可用 **undo gtsm** { **all** | **peer** *ip-address* }命令删除与所有或指定的 LDP 对等体建立 LDP 会话时所配置的 GTSM 功能

4.6.5　LDP GTSM 配置示例

如图 4-14 所示。各节点间运行 MPLS 和 MPLS LDP 协议。为了防止攻击者模拟真实的 LDP 协议单播报文，对 LSRB 发送报文，导致系统异常繁忙，CPU 占用率高的问题。现要求采用 GTSM 功能对节点进行保护，防止非法的 LDP 报文攻击，增强系统安全性。

1. 基本配置思路

通过 GTSM 预防 LDP 报文攻击的方法是通过设置收到对等体设备（可以是直接连接的，也可以是非直接连接的）的 LDP 报文中 TTL 值来实现的。配置的方法也很简单，就是按照 4.6.4 节介绍的表 4-20 步骤进行。当然前提也是先要配置好骨干网各节点的 MPLS 基本能力和 LDP 会话（可以是本地会话，也可以是远端会话）。

由以上分析可以得出本示例如下的基本配置思路。

（1）配置各设备接口 IP 地址和 OSPF 协议，实现骨干网三层互通。

（2）配置各设备 MPLS 基本能力和 LDP 本地会话（假设本示例各节点间仅建立 LDP

本地会话）。

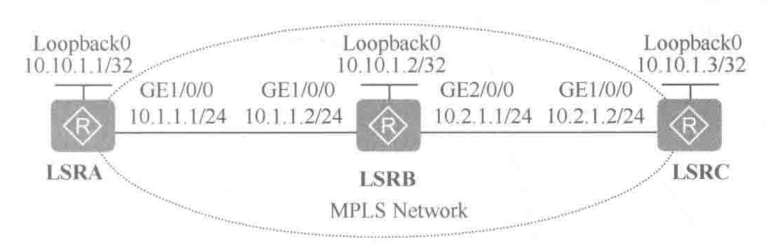


图 4-14　LDP GTSM 配置示例的拓扑结构

（3）配置各设备的 GTSM，每个设备上针对相邻对等体发来 LDP 报文的 TTL 范围可根据实际需要来控制，防止非法设备发来的 LDP 报文与本地设备建立 LDP 会话。

2. 具体配置步骤

（1）配置各设备接口的 IP 地址和 OSPF 协议，实现骨干网三层互通。

LSRA 上的配置。

```
<Huawei> system-view
[Huawei] sysname LSRA
[LSRA] interface loopback 0
[LSRA-LoopBack0] ip address 10.10.1.1 32
[LSRA-LoopBack0] quit
[LSRA] interface gigabitethernet 1/0/0
[LSRA-GigabitEthernet1/0/0] ip address 10.1.1.1 24
[LSRA-GigabitEthernet1/0/0] quit
[LSRA] ospf 1
[LSRA-ospf-1] area 0
[LSRA-ospf-1-area-0.0.0.0] network 10.10.1.1 0.0.0.0
[LSRA-ospf-1-area-0.0.0.0] network 10.1.1.0 0.0.0.255
[LSRA-ospf-1-area-0.0.0.0] quit
[LSRA-ospf-1] quit
```

LSRB 上的配置。

```
<Huawei> system-view
[Huawei] sysname LSRB
[LSRB] interface loopback 0
[LSRB-LoopBack0] ip address 10.10.1.2 32
[LSRB-LoopBack0] quit
[LSRB] interface gigabitethernet 1/0/0
[LSRB-GigabitEthernet1/0/0] ip address 10.1.1.2 24
[LSRB-GigabitEthernet1/0/0] quit
[LSRB] interface gigabitethernet 2/0/0
[LSRB-GigabitEthernet2/0/0] ip address 10.2.1.1 24
[LSRB-GigabitEthernet2/0/0] quit
[LSRB] ospf 1
[LSRB-ospf-1] area 0
[LSRB-ospf-1-area-0.0.0.0] network 10.10.1.2 0.0.0.0
[LSRB-ospf-1-area-0.0.0.0] network 10.1.1.0 0.0.0.255
[LSRB-ospf-1-area-0.0.0.0] network 10.2.1.0 0.0.0.255
[LSRB-ospf-1-area-0.0.0.0] quit
[LSRB-ospf-1] quit
```

LSRC 上的配置。

```
<Huawei> system-view
[Huawei] sysname LSRC
```

```
[LSRC] interface loopback 0
[LSRC-LoopBack0] ip address 10.10.1.3 32
[LSRC-LoopBack0] quit
[LSRC] interface gigabitethernet 1/0/0
[LSRC-GigabitEthernet1/0/0] ip address 10.2.1.2 24
[LSRC-GigabitEthernet1/0/0] quit
[LSRC] ospf 1
[LSRC-ospf-1] area 0
[LSRC-ospf-1-area-0.0.0.0] network 10.10.1.3 0.0.0.0
[LSRC-ospf-1-area-0.0.0.0] network 10.2.1.0 0.0.0.255
[LSRC-ospf-1-area-0.0.0.0] quit
[LSRC-ospf-1] quit
```

以上配置完成后，在各节点上执行 **display ip routing-table** 命令，可以看到相互之间都学到了彼此的路由。

（2）配置各设备 MPLS 基本能力和 LDP 本地会话，建立 LDP LSP。

LSRA 上的配置。

```
[LSRA] mpls lsr-id 10.10.1.1
[LSRA] mpls
[LSRA-mpls] quit
[LSRA] mpls ldp
[LSRA-mpls-ldp] quit
[LSRA] interface gigabitethernet 1/0/0
[LSRA-GigabitEthernet1/0/0] mpls
[LSRA-GigabitEthernet1/0/0] mpls ldp
[LSRA-GigabitEthernet1/0/0] quit
```

LSRB 上的配置。

```
[LSRB] mpls lsr-id 10.10.1.2
[LSRB] mpls
[LSRB-mpls] quit
[LSRB] mpls ldp
[LSRB-mpls-ldp] quit
[LSRB] interface gigabitethernet 1/0/0
[LSRB-GigabitEthernet1/0/0] mpls
[LSRB-GigabitEthernet1/0/0] mpls ldp
[LSRB-GigabitEthernet1/0/0] quit
[LSRB] interface gigabitethernet 2/0/0
[LSRB-GigabitEthernet2/0/0] mpls
[LSRB-GigabitEthernet2/0/0] mpls ldp
[LSRB-GigabitEthernet2/0/0] quit
```

LSRC 上的配置。

```
[LSRC] mpls lsr-id 10.10.1.3
[LSRC] mpls
[LSRC-mpls] quit
[LSRC] mpls ldp
[LSRC-mpls-ldp] quit
[LSRC] interface gigabitethernet 1/0/0
[LSRC-GigabitEthernet1/0/0] mpls
[LSRC-GigabitEthernet1/0/0] mpls ldp
[LSRC-GigabitEthernet1/0/0] quit
```

以上配置完成后，在节点上执行 **display mpls ldp session** 命令，可以看到 LSRA 和 LSRB、LSRB 和 LSRC 之间的本地 LDP 会话状态为“Operational”。

（3）配置各设备的 LDP GTSM。

在这项配置任务中，关键是要确定好每个设备与指定对等间交互 LDP 报文时所限制的 TTL 范围。

本示例中，由于各节点间仅建立了 LDP 本地会话，是直接连接的（相隔仅一跳），所以理论上来说，每台设备上配置的允许对等体发来的 LDP 报文中 GTSR TTL 值均为 1 即可，但为了留有一定网络结构扩展余地，在此配置 GTSR TTL 值均为 3。

LSRA 上的配置。

```
[LSRA] mpls ldp
[LSRA-mpls-ldp] gtsm peer 10.10.1.2 valid-ttl-hops 3   #---配置 LSRB 离 LSRA 的最大跳数为 3
[LSRA-mpls-ldp] quit
```

LSRB 上的配置。

```
[LSRB] mpls ldp
[LSRB-mpls-ldp] gtsm peer 10.10.1.1 valid-ttl-hops 3   #---配置 LSRA 离 LSRB 的最大跳数为 3
[LSRB-mpls-ldp] gtsm peer 10.10.1.3 valid-ttl-hops 3   #---配置 LSRC 离 LSRB 的最大跳数为 3
[LSRB-mpls-ldp] quit
```

LSRC 上的配置。

```
[LSRC] mpls ldp
[LSRC-mpls-ldp] gtsm peer 10.10.1.2 valid-ttl-hops 3   #---配置 LSRB 离 LSRC 的最大跳数为 3
[LSRC-mpls-ldp] quit
```

第5章 MPLS TE基本功能配置与管理



如果把前面各章中介绍的静态 LSP 和 LDP LSP 隧道看作普通 MPLS 隧道的话，那本章所介绍的 MPLS TE（流量工程）就是一种具有部分 QoS 功能的特殊 MPLS 隧道。

MPLS TE 隧道的特殊性主要体现在两个方面：（1）动态 LSP 的建立不是通过 LDP 完成的，而是通过 RSVP-TE（基于流量工程的资源预留协议）完成的，所建立的 LSP 称之为 CR-LSP（基于约束路由的 LSP）。RSVP-TE 不是像 LDP 那样仅依据路由路径来建立 LSP，而是该路径还必须满足一定的约束条件，如路由度量、链路带宽、预留带宽、显式路径等。这样做的目的是使建立后的 LSP 最大限度地满足用户需求，优化各种不同流量的传输路径。（2）在一条 MPLS TE 隧道中可以建立多条 CR-LSP，以实现多 LSP 的负载分担，这在普通的 MPLS 隧道中也是不可以的。但 MPLS TE 隧道与普通 MPLS 隧道一样，既可单独使用，也可应用于各种二层或三层 MPLS VPN 中（有关 MPLS VPN 参见《华为 MPLS VPN 学习指南》）。

本章主要介绍与 MPLS TE 相关的各方面技术原理，以及基本的静态/动态 MPLS TE 隧道配置与管理方法。

5.1 MPLS TE 基础

MPLS TE（MPLS Traffic Engineering，MPLS 流量工程）可通过建立基于一定约束条件的 MPLS 隧道，并将流量引入到一条或多条这样的 MPLS TE 隧道中进行转发，使特定的网络流量按照指定的路径进行传输，达到优化流量传输路径或多条隧道负载进行分担的目的。

说明 在 AR G3 系列路由器中，AR120&AR150&AR160&AR200&AR500&AR510 不支持 MPLS TE 特性。S 系列交换机中仅 S5700 系列及更高端系列支持，但 SI 和 LI 子系列除外。具体参见相应产品手册。

5.1.1 MPLS TE 简介

传统的 IP 网络中，设备通常都是按照到达目的网络的路由路径长短来作为最优路由选择的依据，不考虑路径上的链路带宽等因素，当某条路径发生拥塞时，路由选择协议又都可根据配置将流量切换到其他的备份路径上。这样的选路方式容易出现流量集中于最短路径而导致拥塞，而其他可选链路则较为空闲。

如图 5-1 所示，假设每个链路的 metric（度量，如 OSPF、IS-IS 中 Cost）值相同，且每段链路（即每个相邻设备间的链路）的带宽都是 100Mbit/s。假设，Router_1 向 Router_4 发送的流量为 40Mbit/s，Router_7 向 Router_4 发送的流量为 80Mbit/s。如果 IGP 路由的计算都是基于最短路径优先（SPF）原则的，则所有流量均经过路径 Router_2→Router_3→Router_4 传输，因为这条路径比另一条路径 Router_2→Router_5→Router_6→Router_4 的开销小，此时 Router_2→Router_3→Router_4 路径会出现过载（负载共为 120Mbit/s，超过了链路的物理带宽 100Mbit/s）而引起拥塞，而 Router_2→Router_5→Router_6→

Router_4 路径则处于空闲状态，造成资源浪费。

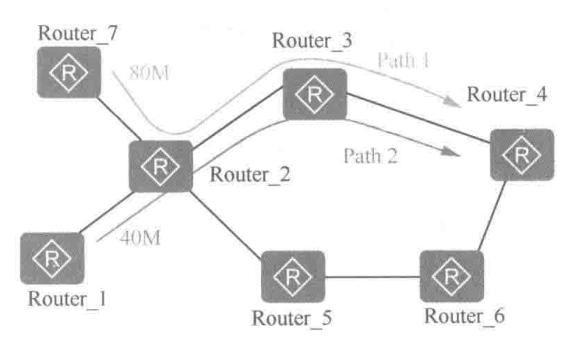


图 5-1　传统路由选路示例

针对这种由于网络资源分配不合理引起的拥塞问题，可以通过流量工程来解决，即将一部分流量分配到空闲的链路上，使网络中流量的分配更加合理。流量工程技术关注的是网络整体性能的优化，以便提供高效、可靠的网络服务，优化网络资源的利用、优化网络流量传输路径的选择。

在 MPLS TE 出现之前，有如下两种流量工程的解决方案。

（1）IP 流量工程

通过调整路径 Metric（如调整接口的 Cost）而控制网络流量的传输路径，这种解决方法能够解决某些链路上的拥塞，但又可能会引起另外一些链路拥塞，因为某条路径上的 Metric 改变后，可能使流量又全部转移到另一条路径上。另外，在拓扑结构复杂的网络上，Metric 值的调整比较困难，往往一条链路的改动会影响多条路由，难以把握和权衡。

（2）ATM 流量工程

现有的 IGP 协议都是拓扑驱动的，只考虑网络的连接情况，不能灵活反映链路带宽和流量特性这类动态状况。解决这种缺陷的一种方法是使用 IP over ATM 重叠模型。然而，实际应用中实施 ATM 流量工程时额外开销大且可扩展性差。

为了在大型骨干网络中部署流量工程，必须采用一种可扩展性好、简单的解决方案。MPLS 可以方便地在物理的网络拓扑上建立一个虚拟的拓扑，然后将流量映射到这个虚拟拓扑上。因此，MPLS 与流量工程相结合的技术应运而生，即 MPLS TE。在 MPLS TE 中建立的是一种特殊的 LSP——CR-LSP（Constraint-based Routed Label Switched Paths，基于约束路由的 LSP）。

CR-LSP 是由 RSVP（Resource Reservation Protocol，资源预留协议）-TE 基于一定约束条件建立的 LSP。与普通 LSP 不同，CR-LSP 的建立不仅依赖路由信息，还需要满足其他一些条件，比如预留带宽需求、显式路径等。所谓“显式路径”就是明确指定的路径，可以明确指定某条 LSP 路径中必须经过的节点设备，也可以明确指定某条 LSP 路径中不能经过的节点设备。而且 CR-LSP 只有唯一的路径，**即使经过计算有各方面条件都一致的多条路径，也必须仅选择其中唯一的一条路径**，这就是涉及路径“仲裁”问题，具体将在本章后面介绍。

对于图 5-1 中的拥塞问题，MPLS TE 可通过建立一条带宽为 80Mbit/s、路径为 Path1

的 LSP，另一条带宽为 40Mbit/s、路径为 Path2 的 LSP，并通过设置将 Router_1 和 Router_7 发送到 Router_4 的流量分别固定引入到这两条 LSP 中传输，就可以而解决上述拥塞问题，如图 5-2 所示。

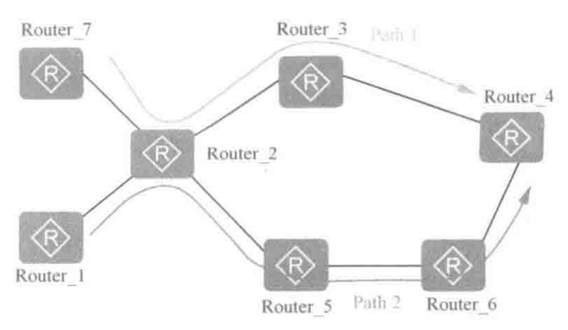


图 5-2 MPLS 流量工程示意

MPLS 技术是在传统的 IP 网络中增加了面向连接的特性，从而使得在传统 IP 网络中实施流量工程成为可能。MPLS TE 技术可以在不进行硬件升级的情况下对现有网络资源进行合理调配和利用，并对网络流量提供带宽和 QoS 保证，最大限度地节省企业成本。同时，MPLS TE 具有丰富的可靠性技术（如 BFD for MPLS TE、RSVP GR、TE FRR 等），能够给骨干网络提供网络级和设备级的可靠性。

5.1.2 RSVP-TE 简介

5.1.1 节已介绍到，在 MPLS TE 隧道中建立 CR-LSP 的信令协议是 RSVP-TE。RSVP-TE 是由传统 IP 网络中的 RSVP 协议（在 RFC 2205 中定义）基于 MPLS TE 隧道扩展而来，最初为 RFC3209，后面依次更新为 RFC3473、RFC 4875、RFC 5420。

RSVP 是一种传输层的控制协议，不参与应用层数据的传送，主要用来通告和维护网络中的保留资源，可以实现路径的建立、维护、拆除和错误通告。RSVP 最初是为了解决 IP 网络中的 QoS 问题，可用它来为数据传输预留带宽，保证其服务质量。由 RSVP 扩展后的 RSVP-TE 可以支持流量工程，可以从多条并行或备选路径中有效地选择一条最佳路径，以平衡网络中不同链路上的业务载荷。

RSVP-TE 协议除继承了 RSVP 协议的带宽预留功能外，还支持了 MPLS 网络中的 MPLS 标签的分发功能。相对 RSVP 来说 RSVP-TE 新增了五个对象：标签请求（LABEL_REQUEST）、标签（LABEL）、显式路径（EXPLICIT_RIUTE）、记录路径（RECORD_ROUTE）和会话属性（SESSION_ATTRIBUTE）。这些新增的对象除了可以携带标签信息外，还可以携带对 LSR 选路时的限制性要求，从而可以建立由入节点到出节点间的 CR-LSP。

RSVP-TE 可以为每条 CR-LSP 预留指定的带宽资源，以确保所建立的每条 CR-LSP 都有一定的带宽保障。RSVP-TE 与 RSVP 一样，只为单向的数据流请求资源，即在双向通信中，需要为双向数据流分别进行资源预留。RSVP-TE 支持以下三种资源预留风格（目前华为设备仅支持 FF 和 SE 这两种）。

（1）FF（Fixed-Filter style，固定过滤风格）

FF 风格是为每个资源预留请求消息（Resv 消息）的发送者（Egress 节点）创建单

独的资源预留，该预留不与其他发送者共享，是属于资源独占模式。同一链路上需要为不同发送者进行不同的资源预留，在一个给定的链路上分配的总带宽预留是为所有请求资源预留的发送者分配的预留带宽总和。但 CR-LSP 与 LDP LSP 一样是点对点的，所以相同发送者与不同接收者（Ingress 节点）之间必须建立不同的 CR-LSP。这样一来也就决定了，如果采用 FF 预留风格，则每个 CR-LSP 有不同的资源预留。FF 预留风格适用于选择同一发送者的不同接收者共享一个节点的预留带宽情形。

（2）SE（Shared Explicit style，共享显式风格）

SE 风格允许为与同一资源预留请求消息的接收者会话、指定的一系列发送者创建共享的资源预留，使同一链路上相同接收者的不同 CR-LSP 共享一个资源预留。该方式主要用于 Make-Before-Break，该机制具体原理将在 5.4.2 节介绍。

SE 风格虽然与下面将要介绍的 WF 风格一样都是共享资源预留方式，但它有一个明确的共享预留资源的发送者范围，而下面的 WF 风格中共享预留资源的发送者范围是没有固定的。

（3）WF（Wildcard-Filter style，通配过滤风格）

WF 风格在链路上建立一个资源预留，允许与同一资源预留请求消息的接收者会话中的所有发送者共享。这种风格对于并不是所有发送者都同时发送流量的组播应用非常有用。例如在一个电话会议的应用中，同时说话的人是有限的。

5.1.3　RSVP-TE 消息类型

RSVP-TE 在 CR-LSP 建立过程中需要用到多种消息，如 Path、Resv、PathTear、ResvTear、PathErr、ResvErr、ResvConf 等。

1. Path 消息

Path 消息是 RSVP-TE 的 CR-LSP 建立请求消息。它是由入节点（Ingress）发送，请求与出节点（Egress）建立预留了带宽资源 CR-LSP。下游路径上除出节点外的其他每个节点都继续向下游节点转发这个消息。

2. Resv 消息

Resv 消息是 RSVP-TE 的资源预留请求消息，是对 Path 消息的应答。它是由出节点发送，携带资源预留信息逆着原来 Path 消息的路径逐跳向上游转发，直到入节点，为入节点请求建立的 CR-LSP 分配标签（对于本地设备来说是入标签，对于上游设备来说是出标签），并沿途向上游节点提出资源预留请求。

3. PathErr 消息

当 RSVP 节点在处理 Path 消息时发生错误，就向上游发送 PathErr 消息，直到入节点。节点在收到 PathErr 消息时不会改变节点状态，只会继续向上游节点转发该 PathErr 消息，最终到达入节点。

4. ResvErr 消息

当 RSVP 节点在处理 Resv 消息时发生错误，就向下游发送 ResvErr 消息，直到出节点。收到 ResvErr 消息的节点不修改本身的状态，只是逐跳转发给下游节点，最终将到达出节点，在出节点上根据错误类型和具体错误内容进行处理。如果错误是由于记录路径导致超过接口的 MTU 引起的，则会向入节点方向发送 PathErr 消息，否则发送 Resv

消息，拆除 CR-LSP，释放资源和预留的标签。

5. PathTear 消息

PathTear 消息由入节点发送，用来沿途删除节点的路径状态信息，直到出节点，其作用与 Path 消息相反。

6. ResvTear 消息

ResvTear 消息由出节点发送，用来沿途删除节点的预留状态，直到入节点，其作用与 Resv 消息相反。

7. ResvConf 消息

ResvConf 消息由入节点发送，向下游逐跳转发，直到出节点，用于对出节点发出的资源预留请求进行确认。仅当出节点发出的 Resv 消息中包含 RESV_CONFIRM 对象时才会发送 ResvConf 消息。

8. Srefresh 消息

Srefresh 消息用来刷新 RSVP 状态，每个节点都可以发送。

5.1.4 RSVP-TE 的对象类型

RSVP-TE 除继承了 RSVP 中的对象外，还新增一些对象。本节对一些主要对象进行简单介绍。

1. SEESION 对象

包含会话的目的节点及相关隧道信息，如源 IP 地址、目的 IP 地址、Tunnel ID，用来定义一个特定的会话。所有 RSVP 消息中都必须包含 SESSION 对象。

2. RSVP_HOP 对象

包含发送 RSVP 消息的前一个节点的 IP 地址和 Tunnel 接口信息，可在多种 RSVP 消息中存在。如果是 Path、PathTear、ResvErr 消息，IP 地址是下游节点的 IP 地址，如果是 Resv、ResvTear、PathErr 消息，IP 地址是上游节点的 IP 地址。

3. TIME_VALUES 对象

定义刷新周期，即发送刷新消息的时间间隔。接收到该包含该对象的消息后，就知道什么时候该发送刷新消息，并计算超时时间。在 Path 消息和 Resv 消息中存在。

4. SENDER_RSPEC 对象

定义了发送端数据流的流量特性，在 Path 消息中存在。

5. SENDER_TEMPLATE 对象

包含了发送端的 IP 地址、LSP ID，与 SEESION 对象一起唯一确定一条 CR-LSP。在 Path 消息中存在。

6. FLOW_SPEC 对象

定义会话数据包的 QoS 特性，在 Resv 消息中存在。

7. FILTER_SPEC 对象

定义了一组会话数据包应当接受的 QoS 特性（在 FLOW_SPEC 对象中定义）。在 Resv 消息中存在。

8. ADSPEC 对象

收集沿途节点是否支持指定的 QoS 参数的信息，用于向下游节点进行通告。在 Path

消息中存在。



以下五个对象是 RSVP-TE 新增的对象。

9. LABEL_REQUEST 对象

用于向下游节点请求分配 MPSL 标签，在 Path 消息中存在。下游节点收到该对象后会保存在自己的 PSB（路径状态块）中，用于在下次转发 Path 消息时包含该对象。但如果收到无法支持的 LABEL_REQUEST 对象，会向上游发送 PathErr 消息，由上游节点逐跳转发，直到入节点。

10. LABEL 对象

用于发布下游节点为上游节点分配的标签。在 Resv 消息中存在。

11. EXPLICIT_ROUTE 对象

简称 ERO，用来标识一个显式路径中的节点。一个 Path 消息可以包括多个该对象，用来指定 CR-LSP 的显式路径。

12. RECORD_ROUTE 对象

简称 RRO，用来记录 CR-LSP 实际经过的路径。当使用的 ERO 丢失时可以使用 RRO 来标记 CR-LSP 路径，可在 Path 消息或 Resv 消息中存在。

13. SESSION_ATTRIBUTE 对象

定义会话的属性，包括 LSP Tunnel 的建立优先级、保持优级、亲和属性和 FRR（快速重路由）等。

5.1.5　RSVP-TE 消息格式

在 5.1.3 节介绍的各类 RSVP 消息都是 IP 报文，对应的协议类型为 46，都包含一个通用的 RSVP 头部，随后是多个可变长度、类型的消息对象。图 5-3 上半部分是 RSVP 消息的通用头部格式，下半部分是通用头部中的“消息对象”部分的格式，各字段的说明见表 5-1。在一个 RSVP 消息中可以携带多个消息对象。

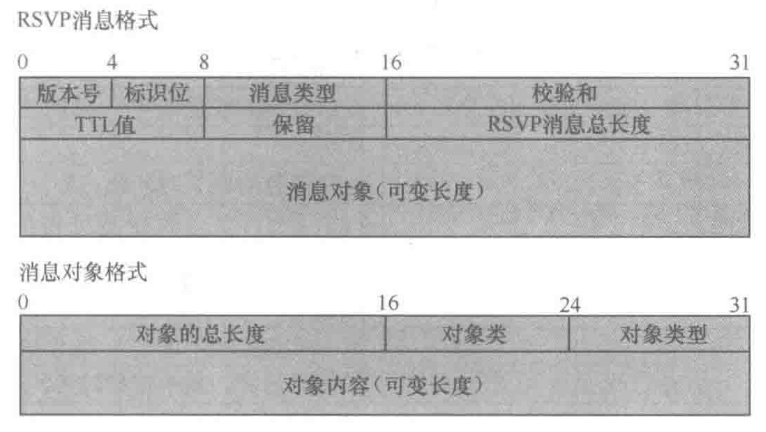


图 5-3　RSVP 消息格式及消息对象格式

表 5-1 **RSVP 消息格式**

字段	长度	描述
以下是 RSVP 通用头中各字段说明，参见图 5-3 上半部分		
版本号	4 比特	表示 RSVP 协议版本号，目前值为 1。
标识位	4 比特	标识位，一般值为 0。RFC2961 扩展其用来标识是否支持摘要刷新（Srefresh）。如果支持 Srefresh，则 Flags 置为 0x01。
消息类型	8 比特	表示 RSVP 消息类型。例如，1 表示 Path 消息，2 表示 Resv 消息。
校验和	16 比特	表示 RSVP 的校验和。如果值为 0，表示消息传输过程中不进行检验和检查，1 为要进行校验和检查。
TTL 值	8 比特	消息的 TTL 值。当节点接收到 RSVP 消息时，通过比较 Send_TTL 和 IP 首部的 TTL 值可以计算出该报文在非 RSVP 域中经过的跳数。
保留	8 比特	保留。
RSVP 消息总长度	16 比特	表示 RSVP 消息的总长度，包括 RSVP 头部和 RSVP 消息，以字节为单位。
消息对象	可变	消息对象，每个 RSVP 消息都包含多个对象。不同类型的消息，包含的对象不同。
以下是“消息对象”部分的各子字段说明，参见图 5-3 下半部分		
对象总长度	16 比特	表示对象的总长度，以字节为单，必须是 4 的倍数，最小值为 4。
对象类	8 比特	对象类。每个对象类都有一个名称，如 SESSION、SENDER_TEMPLATE、TIME_VALUE。
对象类型	8 比特	对象类型，表示同一类对象中不同的类型。Class_Number 与 C-Type 唯一标识了一个对象。
对象内容	可变	对象内容，可变长度

下面介绍在建立 CR-LSP 中使用的 Path 消息和 Resv 消息。

1. Path 消息

在 RSVP-TE 中，Path 消息包括一系列见表 5-2 的对象，用于创建 RSVP 会话和关联路径状态。

表 5-2 **Path 消息包括的对象**

消息对象	对象类（Class）	对象类型（C-type）	对象内容
SESSION	1	1	RSVP 会话相关信息，包括：Destination Address、Tunnel ID、Extend Tunnel ID。
RSVP_HOP	3	1	发送 Path 消息的上一跳的出接口地址和接口索引。
TIME_VALUE	5	1	包含消息的刷新时间值。
SENDER_TEMPLATE	11	1	指定了发送节点的 IP 地址和 LSP ID。
SENDER_TSPEC	12	2	指明了数据流的流量特征。
LABEL_REQUEST	19	1	标签请求对象，只在 Path 消息中携带。
ADSPEC	13	2	用于收集路径上的实际 QoS 相关参数，例如，路径带宽估计、最小路径时延、Path MTU。
EXPLICIT_ROUTE	20	1	ERO（Explicit Route Object，显式路径对象），描述 LSP 经过的路径信息，可以是严格显式路径也可以是松散显式路径。Path 消息沿 ERO 指定的路径转发，不受 IGP 最短路径约束。

（续表）

消息对象	对象类（Class）	对象类型（C-type）	对象内容
RECORD_ROUTE	21	1	RRO（Record Route Object，记录路径对象），Path 消息实际途经的 LSR 的列表。RRO 可用于收集实际的路径信息，发现路由环路，还可以被复制到下一条 Path 消息中以实现路径锁定。
SESSION_ATTRIBUTE	207	1：LSP_TUNNEL_RA 7：LSP Tunnel	指定了建立优先级、保持优先级、资源预留风格、亲和属性等属性，是 RSVP-TE 新增的对象

Path 消息是从 Ingress 节点沿着数据流方向发送到 Egress 节点（源 IP 地址是 Ingress 节点的 LSR ID，目的 IP 地址是 Egress 节点的 LSR ID），途经的节点上会生成 PSB（Path State Block，路径状态块），并启动老化定时器，PSB 用于指导 Resv 消息返回时向上游设备传输。但如果节点收到了一个还没有老化、但收到内容重复的 Path 消息时，PSB 不会更新，本地节点也不会立即向下游设备传输该重复的 Path 消息，而是要等到原来相同内容的 Path 消息的老化定时器超时后才向下游转发该重复的 Path 消息。

2. Resv 消息

当 Egress 节点收到 Path 消息时，将发送一个 Resv 消息进行响应。Resv 消息携带了资源预留信息，从 Egress 节点逐跳发送给上游节点（目的 IP 地址是上游节点的 IP 地址，即每一节点独立发送 Rsev 消息），直到 Ingress 节点。

Resv 消息中包携带有资源预留风格、为 CR-LSP 所分配的标签等内容。沿途的每个节点会生成和维护 RSB（Reserved State Block，源预留状态块）并分配标签，发送完一个 Resv 消息会同时启动一个老化定时器，以便 Resv 消息周期性地发送。RSB 中的标签用于在隧道中传输时指导数据流的转发，类似于 MPLS LDP LSP 标签。当 Resv 消息到达 Ingress 时，一条 CR-LSP 就建立成功了。表 5-3 中列出了包含在 Resv 消息中的一些对象。

表 5-3　Resv 消息包括的对象

消息对象	对象类	对象类型	对象内容
INTEGRITY	4	1	包含 RSVP 消息的认证密钥。
SESSION	1	1	包含了 RSVP 会话相关信息，如 Destination Address、Tunnel ID、Extend Tunnel ID。
RSVP_HOP	3	1	包含发送 Resv 消息出接口 IP 地址和接口索引。
TIME_VALUE	5	1	包含消息的刷新时间值。默认值为 30 秒。
STYLE	8	1	资源预留风格，指 RSVP 节点处理上游节点的资源预留请求时的资源预留方式，由 Ingress 节点指定。华为设备中支持的资源预留风格包括 FF 和 SE，具体参见 5.1.2 节。
FLOW_SPEC	9	1：Reserved (obsolete) flowspec object 2：Inv-serv flowspec object	指明了数据流的 QoS 特征。
FILTER_SPEC	10	1	发送节点的 IP 地址和 LSP ID。
RECORD_ROUTE	21	1	RRO，沿途收集的各节点的入接口 IP 地址、LSR-ID 和出接口 IP 地址。
LABEL	16	1	分配的标签。
RESV_CONFIRM	15	1	预留确认请求，携带了请求预留确认的节点的 IP 地址

5.1.6 MPLTS TE 隧道

以上各章经常提到，除了最常用的 LSP（包括 LDP LSP 和 BGP LSP）隧道外，还有 MPLS TE 隧道和 GRE 隧道。GRE 隧道我们已在《华为 VPN 学习指南》中有详细介绍，但 MPLTS TE 隧道至今仍比较模糊，本节就要专门来介绍这种隧道。

MPLS TE 隧道也是从 Ingress 节点到 Egress 节点的一条虚拟点到点连接，专用于 MPLS TE 中。通常情况下，MPLS TE 隧道由一条 CR-LSP 构成。在部署 CR-LSP 备份或需要将流量通过多条路径传输等情况下，需要为同一种流量建立多条 CR-LSP，此时 MPLS TE 隧道是由一组 CR-LSP 构成。Ingress 节点上 MPLS TE 隧道由 MPLS TE 模式的 Tunnel 接口标识，当流量的出接口为该 Tunnel 接口时，流量将通过构成 MPLS TE 隧道的 CR-LSP 来转发。

说明 在本书前面几章中已说到，一条 LSP 对应一条 MPLS 隧道，那只是针对非 MPLTS TE 隧道而言。在 MPLS TE 隧道中可以包括多条 CR-LSP，但也是单向的。

MPLS TE 隧道涉及以下几个概念。

- 隧道接口：隧道接口，即 TE 模式 Tunnel 接口，是为实现报文的封装而提供的一种点对点类型的虚拟接口，与 Loopback 接口类似，都是一种逻辑接口。
- 隧道标识（Tunnel ID）：采用十进制数字来标识唯一一条 MPLS TE 隧道，以便对隧道进行规划和管理，这个数字称为 Tunnel ID。
- LSP 标识（LSP ID）：采用十进制数字来标识唯一一条 CR-LSP，以便对 CR-LSP 进行规划和管理，这个数字称为 LSP ID。

如图 5-4 所示，有两条 LSP（都是 CR-LSP），其中一条主 LSP（Primary LSP）LSRA→LSRB→LSRC→LSRD→LSRE 作为主路径（假设其 LSP ID=2），另外一条备份 LSP（Backup LSP）LSRA→LSRF→LSRG→LSRH→LSRE 作为备份路径（假设其 LSP ID=1024），而两条 LSP 都对应同一个隧道 ID 为 100 的 MPLS TE 隧道 Tunnel0/0/1。

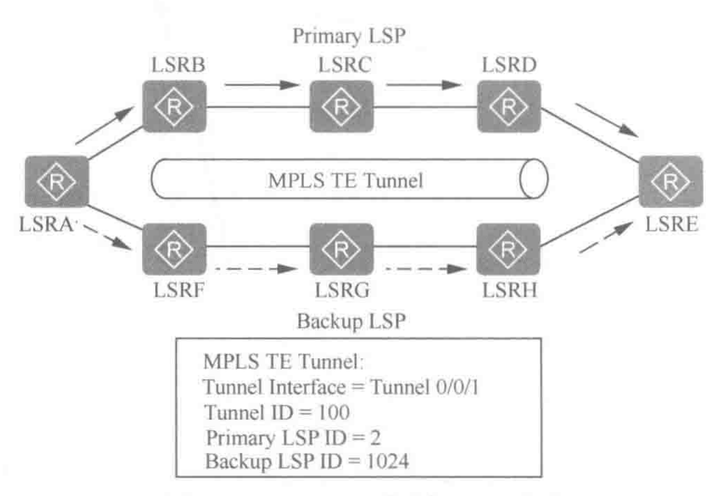


图 5-4 MPLS TE 隧道与 LSP 隧道

5.1.7 MPLS TE链路属性

上文提到，MPLS TE可以依据链路带宽为不同隧道进行流量分配，所以在MPLS TE隧道中还涉及一些链路属性。链路属性是用于标识一条物理链路上的带宽资源使用情况、路由成本及链路的可靠性，具体包括下以几方面的内容。

1. 链路总带宽

物理链路所具有的总带宽值。

2. 最大可预留带宽

本链路中可以预留给MPLS TE隧道使用的带宽值，最大可预留带宽小于等于链路总带宽。如果链路上原来已建立了其他隧道，则可预留带宽还要减去这部分已使用的隧道带宽。

3. 每优先级可用带宽

MPLS TE隧道可配置0～7共8个隧道优先级，优先级值越小，优先级越高。每个优先级的隧道都有一个最大可用带宽（接口会列出各优先级隧道当前最大可用带宽），高优先级隧道可以抢占低优先级隧道的带宽。有关隧道带宽抢占功能将在下节介绍。

如一条100Mbit/s链路，在没有为任何优先级隧道配置预留带宽时，0～7优先级的隧道可预留的带宽均为100Mbit/s。当在这条链路上建立了一条优先级为3的CR-LSP隧道，并且配置为其预留30Mbit/s带宽时，0～2优先级隧道的最大可用带宽仍为100Mbit/s，因为他们的优先级高于优先级3，可以抢占优先级为3的CR-LSP隧道带宽，但是4～7优先级的CR-LSP隧道当前最大可用带宽就只有70Mbit/s了。

假设再在这条链路上了一条优先级为5的CR-LSP隧道，则0～2优先级带宽仍有100Mbit/s的最大可用带宽，优先级4的CR-LSP隧道仍可以有70Mbit/s的最大可用带宽，因为新建的优先级5的CR-LSP隧道不会抢占高优先级的带宽，但6～7优先级的CR-LSP隧道就只有40Mbit/s了，因为要减去优先级为3、5的两条隧道所配置的带宽。

4. TE Metric

链路的TE度量。为了增强对TE隧道路径计算中的可控性，MPLS TE提供了TE Metric，使得隧道在计算路径时能更独立于IGP的路由选路。缺省情况下，采用IGP的度量值作为TE度量值，如采用OSPF、IS-IS路由的链路开销（Cost）。

5. SRLG

SRLG（Shared Risk Link Group，共享风险链路组）是一组共享同一个公共物理资源（如共享一根电缆、一根光纤）的多条MPLS TE链路。同一个SRLG的链路具有相同的风险等级，即如果SRLG中的一条MPLS TE链路失效，组内的其他链路也失效。这主要是为了保证MPLS TE隧道的带宽或CR-LSP热备份。

SRLG主要用在CR-LSP热备份和TE FRR组网中增强TE隧道的可靠性。

6. 链路管理组

也称为链路颜色，或链路属性，是一个表示链路属性的32位向量，每一个比特位可以表示链路的一个属性，如链路带宽、性能或者管理策略（比如标识这段链路上有MPLS TE隧道经过，或者这段链路上承载的为组播业务）。链路管理组需要和下节将要介绍的MPLS TE隧道属性中的“亲和属性”配合使用来达到控制隧道路径的目的。

5.1.8 MPLS TE 隧道属性

MPLS TE 隧道所使用的 CR-LSP 是基于一定约束条件建立的 LSP，这些约束条件就称之为隧道属性。这些约束条件包括带宽约束和路径约束两个方面。

（1）带宽约束

为隧道预留的最大可用带宽，是根据每个隧道优先级进行配置的。

（2）路径约束

主要包括显式路径、优先级与抢占、路径锁定、亲和属性和跳数限制。

建立和管理约束条件的机制称为 CR（Constraint-based Routing，基于约束的路由），下面对 CR 的主要内容进行介绍。

1. 隧道带宽

隧道的带宽值需要根据隧道要承载的业务进行规划，隧道建立时将根据这个值要求隧道沿途的节点进行相应带宽预留，从而为特定隧道中的业务提供带宽保证。

2. 显式路径

显式路径是指在 CR-LSP 建立时由用户手工明确指定其必须经过或避开指定节点的路径，其指定方式又分为以下两种。

（1）严格显式路径

严格方式可以指定路径上必须经过哪些节点，下一跳与前一跳必须直接相连。通过严格显式路径，可以最精确地控制 CR-LSP 所经过的路径。

如图 5-5 所示，LSRA 作为 CR-LSP 的 Ingress 节点，LSRF 作为 Egress 节点，从 LSRA 到 LSRF 用严格显式路径建立一条 CR-LSP：LSRA→LSRB→LSRC→LSRE→LSRD→LSRF。“LSRB Strict”表示该 CR-LSP 必须经过 LSRB，并且 LSRB 的前一跳是 LSRA，“LSRC Strict”表示该 CR-LSP 必须经过 LSRC，并且 LSRC 的前一跳是 LSRB，依此类推，就可以精确控制该条 CR-LSP 所经过的路径。

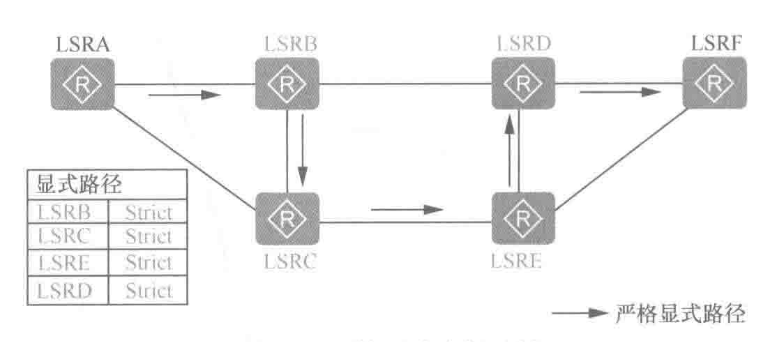


图 5-5 严格显式路径示例

（2）松散显式路径

松散方式可以指定路径上必须经过哪些节点，而且列出的节点之间可以有其他节点。如图 5-6 所示，从 Ingress 节点 LSRA 到 Egress 节点 LSRF 用松散显式路径建立一条 CR-LSP。“LSRD Loose”表示该 CR-LSP 必须经过 LSRD，但是从 LSRA 到达 LSRD 可以有多条路径，如可以经过 LSRB 直接到达；也可以经过 LSRC，再经过 LSRE 到达。

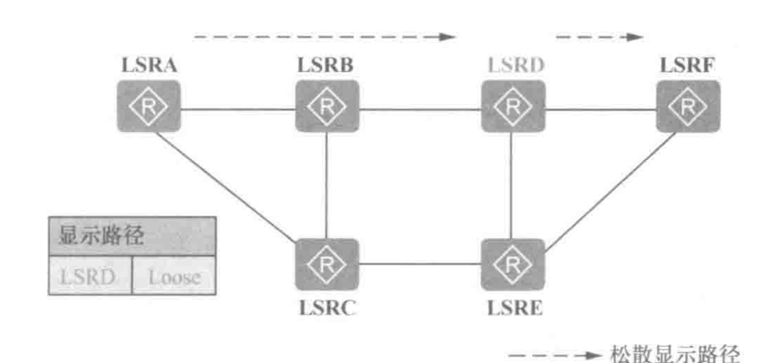


图5-6 松散显式路径示例

3. 隧道优先级与带宽抢占

隧道优先级与带宽抢占功能可根据MPLS TE隧道承载业务的重要程度来解决隧道建立过程中的资源竞争问题。

优先级用来指示新建路径和当前已经建立路径之间的带宽资源的抢占关系。MPLS隧TE道使用“建立优先级”（Setup Priority）和“保持优先级”（Holding Priority）来决定是否可以进行带宽资源的抢占。所谓“建立优先级”是指建立TE隧道时使用带宽资源的优先级，而“保持优先级”是指保持当前TE隧道所使用的带宽资源的优先级。

如果新建TE隧道的“建立优先级”高于已经建立TE隧道的“保持优先级”，则新建TE隧道会抢占已有TE隧道的带宽资源，反之则不会发生带宽资源抢占。建立优先级和保持优先级的范围都是从0到7，**7为最低优先权**（与QoS优先级相反）。**同一条隧道建立优先级不能高于保持优先级**。

如果在建立TE隧道的过程中，当建立高优先级的TE隧道无法找到满足所需带宽要求的路径时，则拆除当前已经建立的另一条TE隧道，占用为它分配的带宽资源，这种处理方式称为抢占（Preemption）。抢占分为硬抢占和软抢占两种。

- 硬抢占：高优先级的TE隧道和低优先级TE隧道发生资源竞争时，高优先级TE隧道直接抢占低优先级TE隧道的资源。即当链路总带宽不够两TE隧道共享时，直接拆除原来低优先级TE隧道，否则两条TE隧道共享链路带宽。
- 软抢占：高优先级的TE隧道和低优先级TE隧道发生资源竞争，且链路总带宽不够两条TE隧道共享时，采用Make-Before-Break的原则，即先不拆除低优先级TE隧道，而是等到低优先级TE隧道的流量切换到新的TE隧道后，高优先级TE隧道才抢占原低优先级TE隧道的资源，并拆除原来低优先级的TE隧道，否则两条TE隧道共享链路带宽。

如图5-7所示，各个链路的带宽分布如图中所示，每个链路的Metric值都相同。假设已存在两条TE隧道。

- Tunnel 0/0/1：路径为Path1，带宽需求为100Mbit/s，建立和保持优先级为0。
- Tunnel 0/0/2：路径为Path2，带宽需求为100Mbit/s，建立和保持优先级为7。

现如果LSRB→LSRE之间的链路发生了故障，则Tunnel0/0/1原来对应的Path1路径不通了，此时LSRA会重新计算出Tunnel 0/0/1新路径Path3（LSRA→LSRB→LSRF→LSRE）。但LSRB→LSRF这段链路的带宽仅为100Mbit/s，不够Tunnel 0/0/1、Tunnel 0/0/2两条隧道共同使用，于是将发生两条隧道对这段链路的带宽资源抢占，如图5-8所示。

Tunnel 0/0/1 的新路径建立过程如下。

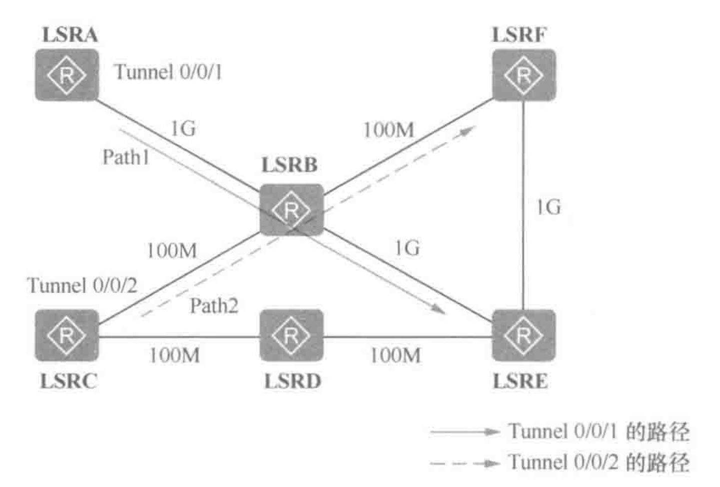


图 5-7 隧道优先级与抢占示例——链路发生故障前的链路状态

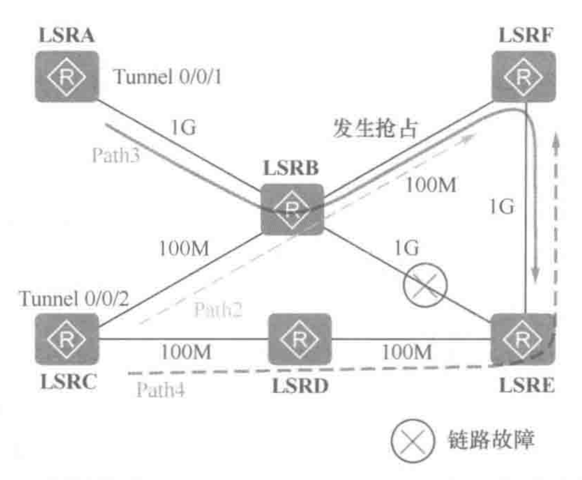


图 5-8 隧道优先级与抢占示例——发生抢占时的链路状态

（1）经过 MPLS TE 的路径计算后，Path 消息沿着 LSRA→LSRB→LSRF→LSRE 发送，Resv 消息沿着 LSRE→LSRF→LSRB→LSRA 发送。

说明 Path 消息是由发送者向下游转发，保存所经过节点的路径信息。Resv 消息是由接收者向上游逐跳转发，用于响应 Path 消息，提出资源预留请求。具体参见 5.1.3 节介绍。

（2）当 Resv 消息从 LSRF 发送到 LSRB，要在 LSRB 为新建 TE 隧道进行带宽预留时，发现带宽不足（因为 LSRB 到 LSRF 之间的链路总带宽仅为 100Mbit/s，不够两条 TE 隧道共享），于是发生抢占。此时对于所采取的不同抢占方式所对应的结果如下。

■ 在硬抢占方式下：由于 Tunnel 0/0/1 的优先级高于 Tunnel 0/0/2，LSRB 将直接拆除 Tunnel 0/0/2 的路径 Path2，并往 LSRF 发 PathTear 消息告知删除节点路径信息，往 LSRC 发送 ResvTear 消息告知删除节点预留状态。此时如果 Tunnel 0/0/2 存在流量，需通过 LSRF 到达目的地的那部分流量将丢失。

说明 PathTear 消息是用来删除节点路径信息，其作用与 Path 消息相反。ResvTear 消息是用来删除节点预留状态，其作用与 Resv 消息相反。具体参见 5.1.3 节介绍。

■ 在软抢占方式下：LSRB 往 LSRC 发送 ResvTear 消息，并在 LSRB 和 LSRC 不拆除 Path2 的前提下，沿着 Path4 重新建立路径，即 LSRC→LSRD→LSRE→LSRF。等路径建立完成且将流量切换后，拆除原有的 Tunnel 0/0/2 的路径 Path2。

【经验提示】在图 5-7 中如果发生故障的不是 LSRB 到 LSRE 这段链路，而是 LSRB 到 LSRF 这段链路。假设 Path2 经过计算，也要通过 LSRB 直接到达 LSRE，这时 Tunnel 0/0/1 隧道和 Tunnel 0/0/2 隧道是否会共享 LSRB 到 LSRE 这段链路的带宽资源呢？答案是会的，因 LSRB 到 LSRE 这段链路一共有 1Gbit/s 的带宽，完全可以供这两条 100Mbit/s 的 MPLS TE 隧道共享，不会出现资源抢占。如果 Tunnel 0/0/2 隧道的优先级高于 Tunnel 0/0/1 隧道优先级，Tunnel 0/0/2 隧道需要与 Tunnel 0/0/1 隧道共享同一链路时，两条 100Mbit/s 的 Tunnel 0/0/1、Tunnel 0/0/2 隧道更会共享是 LSRB 到 LSRE 这段链路的 1Gbit/s 带宽了。

4. 路径锁定

当一条 CR-LSP 建立完成后，网络拓扑变化或者改变某些隧道的属性时，可能导致这条 CR-LSP 根据实时网络状态重新建立。这里存在两个问题。

■ 新建立的 LSP 路径可能与原路径不同，这就不便于网络管理员的运维管理。

■ 流量需要从老的 CR-LSP 切换到新 CR-LSP，这有可能导致流量丢失。

路径锁定功能就可以用来解决上述两个问题。路径锁定是指在 CR-LSP 建立完成后强制其路径不随路由变化而变化，从而使得业务流量具有连续性，并能够提供一定的可靠性保证。但这也可能因为原路径已出现了故障，导致流量的丢失。

5. 亲和属性

亲和属性（Affinity attribute）是描述 MPLS TE 隧道所需链路的 32 位向量值，在隧道的 Ingress 节点来配置实施，需要和前面介绍的“链路管理组”联合使用，供隧道在计算路径时决定该隧道可以使用哪些链路。

亲和属性是由 32 位的属性值和 32 位的掩码组成，与 IPv4 地址和子网掩码的组合类似。每一位代表一个属性，**掩码为 0 时表示不关心对应亲和属性位的值，掩码为 1 时表示关心对应亲和属性位的值**。

MPLS TE 隧道的亲和属性和掩码共同决定了隧道关心哪些属性，就像 IPv4 地址与子网掩码共同确定 IPv4 地址所属的网络一样。然后，再通过 Tunnel 接口上配置的链路管理组属性值与亲和属性匹配，与亲和属性一致的链路管理组属性对应的链路将被选中，从而影响 CR-LSP 的链路选择。

链路管理组属性与亲和属性匹配的方法是把链路管理组属性和亲和属性分别与掩码进行逻辑“与”运算，如果得到的结果相同，则隧道选路时选择该路径，不同则放弃该路径。根据以上匹配规则可以得出链路管理组属性中各位赋值的要求如下。

■ 在所有掩码中为 1 的位中，链路管理组属性中至少有 1 位与亲和属性中的相应位都为 1，否则可能导致链路管理组属性与掩码进行逻辑“与”运算后结果为全 0。**亲和**

属性为 0 的位对应的链路管理组属性位不能为 1，否则可能会导致两者与掩码进行逻辑“与”运算后不一致。

如亲和属性为 0x0000FFFF，掩码为 0xFFFFFFFF，则可用链路的管理组属性取值如下。

- 高 16 位只能取 0，因为本示例中亲和属性的高 16 位全为 0，而规则规定亲和属性为 0 的位对应的链路管理组属性位不能为 1。
- 低 16 位至少有 1 位为 1，因为掩码和亲和属性的低 16 位都全为 1，而规则中规定所有掩码为 1 的位中，链路管理组属性中至少有 1 位与亲和属性中的相应位都为 1。

由此可得出本示例中可使用的链路管理组属性取值范围是 0x00000001～0x0000FFFF。

■ **对于掩码为 0 的位**，则不对链路管理组属性的相应位进行检查，**即对应的链路管理组属性位的值可以任意**。

如亲和属性为 0xFFFFFFFF，掩码为 0xFFFF0000，则可用链路的管理组属性取值如下。

- 高 16 位至少有 1 位为 1，因为掩码和亲和属性的高 16 位均全为 1，而上一条规则规定所有掩码为 1 的位中，链路管理组属性中至少有 1 位与亲和属性中的相应位都为 1。
- 低 16 位则可以任意取 0 或 1，因为掩码的低 16 位全为 0，而规则规定对于掩码为 0 的位，不对链路管理组属性的相应位进行检查，即可以任意是 0 或 1。

由此可得出本示例中可用链路的管理组属性取值范围是 0x00010000～0xFFFFFFFF。

结合以上两项规则，假设亲和属性为 0xFFFFFFF0，掩码为 0x0000FFFF，则可用链路的管理组属性高 16 位可以任意取 0 或 1，17～28 位中至少有 1 位为 1，且低 4 位不能为 1。大家自己计算试一下。

说明 不同设备制造商实现的管理组和亲和属性的比较规则可能有所不同，当在同一网络中使用不同设备制造商的设备时，需要事先了解各自的实现方式，以保证不同制造商设备间能够互通。

缺省情况下，链路管理组属性、亲和属性和掩码值均全为 0，对链路选择不作任何限制。

对于部署 MPLS TE 隧道的网络管理员来说，可以将链路管理组和隧道亲和属性配合使用来达到控制隧道路径的目的。具体配置方法将在第 6 章介绍。

6. 跳数限制

跳数限制值作为 CR-LSP 建立时的选路条件之一，就像链路管理组属性和亲和属性一样，可以限制一条 CR-LSP 允许选择的路径跳数不超过某个值。

5.1.9 MPLS TE 框架

动态 MPLS TE 隧道建立的实现过程主要依靠图 5-9 中的四大功能(也是四大步骤)：（1）通过扩展 IGP（如 OSPF TE 或 IS-IS TE）进行信息发布、收集 TE 相关信息；（2）根据 MPLS TE 信息进行路径计算；（3）然后使用 RSVP-TE 信令协议与上下游节点

交互协议报文来进行路径建立，实现 MPLS TE 隧道的建立；（4）最后将数据报文引入到 MPLS TE 隧道中进行转发。表 5-4 中对四个功能进行了具体的描述。

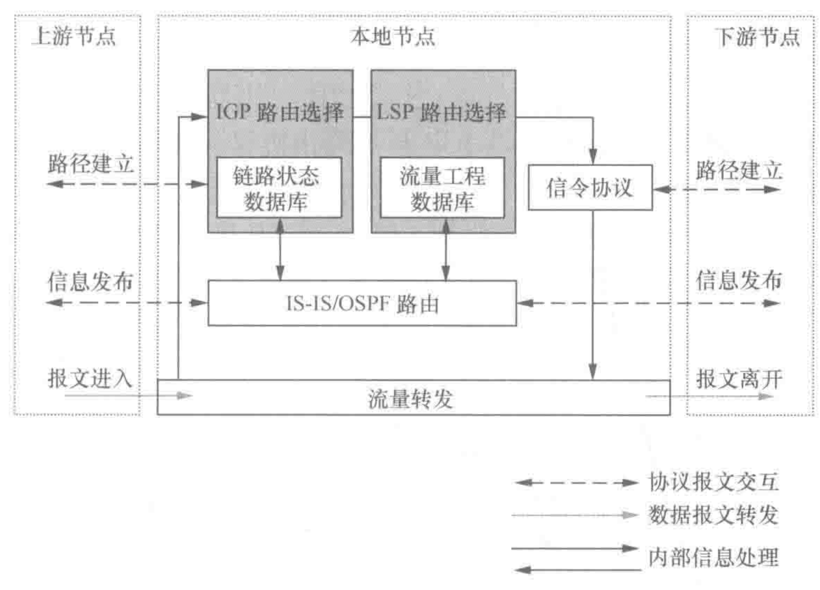


图 5-9　MPLS TE 实现框架

表 5-4　**MPLS TE 实现的四大功能**

序号	功能	功能描述
1	信息发布	除了由 IGP 自己发布的网络拓扑信息外，流量工程还需要知道网络的负载信息。为此，MPLS TE 通过对现有的 IGP 进行扩展，来发布 TE 信息，包括最大链路带宽、最大可预留带宽、当前预留带宽、链路颜色等。每个节点收集本区域所有节点、每条链路的 TE 相关信息，最终生成自己的流量工程数据库 TEDB（TE DataBase）。
2	路径计算	通过 CSPF（Constrained Shortest Path First，约束最短路径优先）算法，利用 TEDB 中的数据来计算满足指定约束条件的路径。CSPF 算法由最短路径优先（SPF）算法演变而来，它首先在当前拓扑结构中删除不满足隧道约束条件的节点和链路，然后再通过 SPF 算法来计算 CR-LSP 路径。
3	路径建立	建立 CR-LSP，CR-LSP 包括以下两种。 ● 静态 CR-LSP 静态 CR-LSP 通过手工配置转发信息和资源信息，不涉及信令协议和路径计算，即不涉及上面的信息发布和路经计算这两个步骤。由于不需要交互 MPLS 相关控制报文，消耗资源比较小，但静态 CR-LSP 不能根据网络的变化动态调整，通常适用于拓扑简单、规模小的组网。 ● 动态 CR-LSP 设备采用 RSVP-TE 信令建立 CR-LSP 隧道，包括本表中所介绍的全部四大功能。RSVP-TE 信令能够携带隧道带宽、显式路径、亲和属性等约束条件。通过信令协议动态地建立 LSP 隧道可以避免逐跳配置的麻烦，适用于规模大的组网。 为了增强路径建立的安全性和可靠性，还可以通过RSVP 认证机制实现。
4	流量转发	将流量引入到 MPLS TE 隧道，并进行 MPLS 转发。前面三个功能可以实现一条 MPLS TE 隧道建立，流量转发用于将进入设备的流量引入到 MPLS-TE 隧道中进行转发

以下 5.2～5.5 节分别介绍以上四大功能的具体工作原理。

5.2 MPLS TE 信息发布原理

MPLS TE 中的信息发布是指通过 IGP 扩展路由协议（通常是采用链路状态类型的 OSPF 或 IS-IS 进行扩展，形成对应的 OSPF TE 或 IS-IS TE）发布网络中各节点的资源分配情况。每台设备收集本区域或本级别（如 IS-IS TE 中的 Level-1、Level-2、Level-1-2）所有设备上每条链路的 TE 相关信息，生成 TEDB（流量工程数据库）。MPLS TE 网络中的各个节点，尤其是隧道的 Ingress 节点将根据信息发布的结果决定隧道经过哪些节点。

5.2.1 MPLS TE 信息内容

通过 OSPF TE 或 IS-IS TE 发布的 MPLS TE 信息的内容主要有以下几种。

■ 链路状态信息：IGP 协议本身收集的信息，如接口 IP 地址、链路类型、链路开销。

■ 带宽信息：包括链路最大物理带宽、最大可预留带宽和每个优先级对应的当前可用带宽，即每个优先级的未被预留带宽。

■ TE Metric：链路的 TE 度量值。缺省情况下，采用 IGP 的度量（如链路开销）值作为 TE 度量值。

■ 链路管理组：链路颜色。

■ 亲和属性：MPLS TE 所需的链路颜色。

■ SRLG：共享风险链路组，作为备份隧道路径计算的限制条件，使得备份路径不与隧道主路径建立在具有同等风险等级的链路上。

MPLS TE 信息的发布主要是依靠现有链路状态路由协议的扩展，包括 OSPF TE 和 IS-IS TE，用于传送带有流量参数的 LSA，满足 MPLS 流量工程的需求。两种 IGP 路由协议会自动收集信息发布内容，并对这些信息进行泛洪，发布给 MPLS TE 网络中的其他节点。

5.2.2 OSPF TE

OSPF 是一种基于链路状态信息的路由协议，具有较强的扩展功能。OSPF 定义了第 1～5、7 类的 LSA 来携带区域内、区域间、自治系统外部等路由信息，并用于路由计算。但是这几种 LSA 的固定格式不能够适应 MPLS TE 的需求，因而出现了 Opaque（透明）LSA 和 TE LSA。

1. Opaque LSA

Opaque LSA 在 RFC2370 中定义，分为三类，分别为第 9、10、11 类 LSA。第 9 类 LSA 只能在某一个接口上扩散，属于链路本地性质的 LSA；第 10 类 LSA 只能在某一个区域内扩散；而第 11 类 LSA 则与第 5 类 LSA 具有相同的扩散范围，可以在除了 Stub（末梢）区域、NSSA（非纯末梢区域）之外的整个自治系统内部扩散。有关 Stub、NSSA 区域的详细技术原理参见《华为路由器学习指南》。

Opaque LSA 与其他几类 LSA 具有相同的头部结构，只是 LSA 中的四字节 Link State ID（链路状态 ID）字段被分为了两部分："Opaque Type"（1 字节）和 "Opaque ID"（3 字节），如图 5-10 所示。

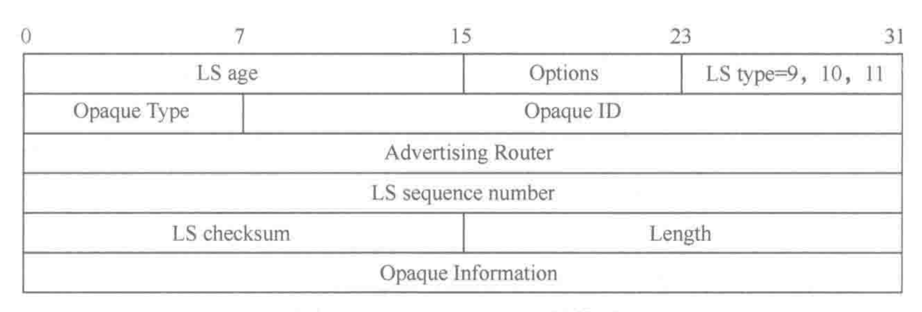


图 5-10　Opaque LSA 的格式

"Opaque Type" 字段用来区分此 LSA 的应用类型，"Opaque ID" 字段用来区分同一种应用类型的不同的 LSA。例如，应用于 OSPF Graceful Restart（平滑重启）的 Opaque LSA 属于第 9 类 LSA，其应用类型为 "3"；而应用于 MPLS TE 扩展 Opaque LSA 属于第 10 类 LSA，其应用类型（Opaque Type 字段）值为 "1"。"Opaque Information" 字段中是 LSA 携带的具体信息，信息格式可由不同的应用根据各自的需求来单独定义。通常采用的格式是一种非常具有扩展能力的 TLV（Type/Length/Value）结构，如图 5-11 所示。

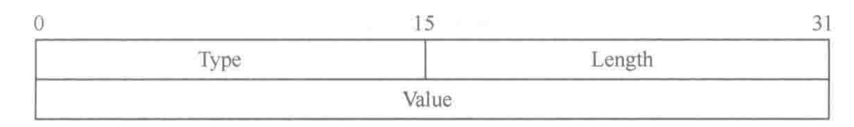


图 5-11　"Opaque Information" 字段的 TLV 结构

- Type：标志了这个结构中携带的信息类型。
- Length：标明了 "Value" 字段的有效字节长度。
- Value：TLV 携带的信息，又是一个 TLV 结构，称为 "子 TLV（sub-TLV）"。

2. TE LSA

以上所说的用于 MPLS TE 扩展的 Opaque LSA 又称为 TE LSA，属于第 10 类 LSA，其应用类型（Opaque Type 字段）值为 "1"。因此 TE LSA 都具有 "1.x.x.x" 形式的 Link State ID，只能在一个区域内扩散。TE LSA 的结构如图 5-12 所示。

TE LSA 使用 TLV 结构来携带需要的信息。目前只定义了两种 TLV。

- TLV Type 1

Router Address TLV：TLV 值为路由器 IP 地址，长度为 4 个字节，用来唯一标识一个 MPLS 节点，在 CSPF 中相当于 OSPF 中的 Router ID 的作用。

- TLV Type 2

Link TLV，携带了使能 MPLS TE 的一条链路的属性。其中，Link TLV 又可携带 9 种 Sub-TLV，具体见表 5-5。OSPF TE 就是通过 TE Type2 LSA 向邻居设备发布最终构建 CR-LSP 所需的 MPLS TE 信息的。

0		15 23	31
LS age		Options	LS type=10
Opq Type=1	Opaque ID		
Advertising Router			
LS sequence number			
LS checksum		length=132	
TLV Type=1		TLV length=4	
Router Address			
TLV Type=2		TLV length=100	
Sub-TLV Type=1		Sub-TLV length=1	
Link Type=1	Padding		
Sub-TLV Type=2		Sub-TLV length=4	
External Route Tag			
Link ID			
Sub-TLV Type=3		Sub-TLV length=4N	
Local IP Address			
Sub-TLV Type=4		Sub-TLV length=4N	
Remote IP Address			
Sub-TLV Type=5		Sub-TLV length=4	
TE Metric			
Sub-TLV Type=6		Sub-TLV length=4	
Maximum Bandwidth			
Sub-TLV Type=7		Sub-TLV length=4	
Maximum Reservable Bandwidth			
Sub-TLV Type=8		Sub-TLV length=32	
Unreserved Bandwidth-Priority 0			
Unreserved Bandwidth-Priority 1			
...			
Unreserved Bandwidth-Priority 7			
Sub-TLV Type=9		Sub-TLV length=4	
Administrative Group			

图 5-12 TE LSA 结构

表 5-5 Type2 LSA 中支持的 Sub-TLV 类型

Sub-TLV	说明
Type1：Link Type（Value 域长度为 1 字节）	链路类型，包括以下两种。 • Point-to-Point：点对点，值为 1。 • MultiAccess：多路访问，值为 2。 因为该子 TLV 仅用来标识链路类型，无具体的值，故该 sub-TLV 的 Value 域后面是 3 字节的填充（Padding）域，填充值全为 0

（续表）

Sub-TLV	说明
Type2：Link ID（Value域长度为4字节）	链路ID，也是外部路由标记（Tag），IP地址格式，对应也有以下两种。 ● Point-to-Point链路：邻居的OSPF Router ID。 ● MultiAccess链路：DR节点的接口IP地址
Type3：Local IP Address（Value域长度为4N字节）	本地接口的IP地址，可以包含多个本地接口的IP地址，每个4字节
Type4：Remote IP Address（Value域长度为4N字节）	对端接口的IP地址，可以包含多个对端接口的IP地址，每个4字节。 ● Point-to-Point链路：使用对端IP地址。 ● MultiAccess链路：可以使用0.0.0.0，也可以省略此sub-TLV
Type5：Traffic Engineering Metric（Value域长度为4字节）	在TE链路上配置的TE Metric。ULONG数据格式，4字节无符号整数
Type6：Maximum Bandwidth（Value域长度为4字节）	链路上的最大带宽。以四个字节存储的浮点数据格式
Type7：Maximum Reservable Bandwidth（Value域长度为4字节）	链路上的最大可预留带宽。以四个字节存储的浮点数据格式
Type8：Unreserved Bandwidth（Value域长度为32字节）	每个优先级（共8个）的可预留带宽。每个优先级都是以四个字节存储的浮点数据格式
Type9：Administrative Group（Value域长度为4字节）	链路管理组属性

当一条链路标记为MPLS TE链路时，若该链路也同时运行了OSPF协议、并且已经建立了OSPF邻居，那么OSPF的TE扩展功能就会根据这一条TE链路产生一条对应的TE LSA发布到区域中。如果区域中有其他的节点也支持TE的扩展，那么在这些节点之间就会产生一个TE链路组成的网络拓扑。每一个发布TE LSA的节点必须具有一个唯一的Router Address。

Opaque Type 10 LSA是在OSPF区域内发布的，所以CSPF计算也是基于区域的，跨区域的LSP需要分段计算。

5.2.3 IS-IS TE

IS-IS也是基于链路状态信息的路由协议，因此也可以使用扩展的IS-IS发布TE信息。在IS-IS TE中扩展了两种新的TLV。

■ Type 135：Wide Metric

IS-IS的有两种度量：（1）Narrow Metric：6比特窄域度量；（2）Wide Metric：32比特广域度量，不用于路由计算，仅用于传递TE相关信息。Narrow Metric只有64（0～63）个度量值，难以满足大型流量工程的需求。因此通常使用Wide Metric来传递TE相关信息。在窄域向广域过渡中，IS-IS TE需要支持以下兼容的度量值。

- Compatible：可以接收和发送度量类型为narrow和wide的报文。
- Wide Compatible：可以接收度量类型为narrow和wide的报文，但只发送wide的报文。

■ Type 22：IS 可达性 TLV

Type 类型为 22 的 IS 可达性 TLV 格式如图 5-13 所示，包括以下字段：

- 系统 ID 和伪节点 ID（System ID and preudonode number），7 个字节。
- 缺省链路度量（Link Metric），5 个字节。
- 子 TLV 长度（sub-TLV Length），1 个字节。
- 可变长的子 TLV（sub-TLVs），0～224 个字节。

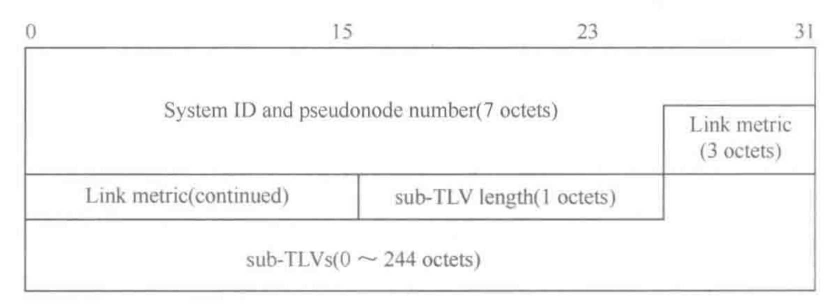


图 5-13 IS 可达性 TLV 格式

IS 可达性 TLV 支持的子 TLV 见表 5-6。IS-IS TE 就是通过这些类型子 TLV 向邻居设备发布最终构建 CR-LSP 所需的 MPLS TE 信息的。

表 5-6 IS 可达性 TLV 的子 TLV 类型

Sub-TLV	描述
Type3：Administrative group（4 字节）	链路管理组属性，以 32 个比特标识 32 个链路管理组属性
Type6：IPv4 interface address（4N 字节）	本地接口的 IP 地址，可以包含多个本地接口的 IP 地址，每个 4 字节
Type8：IPv4 neighbor address（4N 字节）	对端接口的 IP 地址可以包含多个对端接口的 IP 地址，每个 4 字节。 • Point-to-Point 链路：使用对端 IP 地址。 • MultiAccess 链路：使用 0.0.0.0
Type9：Maximum link bandwidth（4 字节）	链路上的最大带宽
Type10：Reservable link bandwidth（4 字节）	链路上的最大可预留带宽
Type11：Unreserved bandwidth（32 字节）	每个优先级（共 8 个）的可预留带宽，每个优先级可预留带宽 4 个字节
Type18：TE Default metric（3 字节）	在 TE 链路上配置的 TE Metric，缺省与 IGP 的度量一致

5.2.4 MPLS TE 信息发布

为了形成本区域内统一的流量工程数据库，OSPF TE 和 IS-IS TE 需要对链路信息进行泛洪。除了首次配置 MPLS TE 隧道会触发泛洪之外，其他的泛洪时机和条件如下。

■ 周期性泛洪（IGP 方式）。

■ 链路生效或失效（IGP 方式）。

■ 链路配置发生变化时，如链路 Cost 发生更新（IGP 方式）。

■ 由于没有足够的资源来预留带宽导致 LSP 无法建立时，该节点会马上泛洪，通告链路的当前可用带宽（TE 特有）。

■ 链路属性发生变化，如链路的管理组和亲和属性发生变化（TE 特有）。

■ 链路带宽发生变化（TE 特有）。

当 MPLS 接口的剩余带宽发生变化时，系统会更新 TEDB 并进行泛洪。当节点上创建大量需要预留带宽的隧道时，系统会频繁更新 TEDB 并泛洪。例如某条链路带宽为 100Mbit/s，在此链路上建立 100 条 1Mbit/s 的 TE 隧道时，则需要进行 100 次泛洪。

为了抑制更新 TEDB 和泛洪的频率，提供了如下带宽泛洪机制：

■ 一条链路上为 MPLS TE 隧道保留的带宽与 TEDB 中的链路剩余带宽的比值等于或大于设定的阈值。

■ MPLS TE 隧道释放的带宽与 TEDB 中剩余带宽的比值等于或大于设定的阈值。

当满足以上两种条件的任意一个时，IGP 将对该链路信息进行泛洪，随之更新 TEDB。例如某条链路剩余带宽为 100Mbit/s，在此链路上建立 100 条 1Mbit/s 的 TE 隧道时，如果设置泛洪阈值为 10%，则变化带宽与剩余带宽的比值如图 5-14 所示。

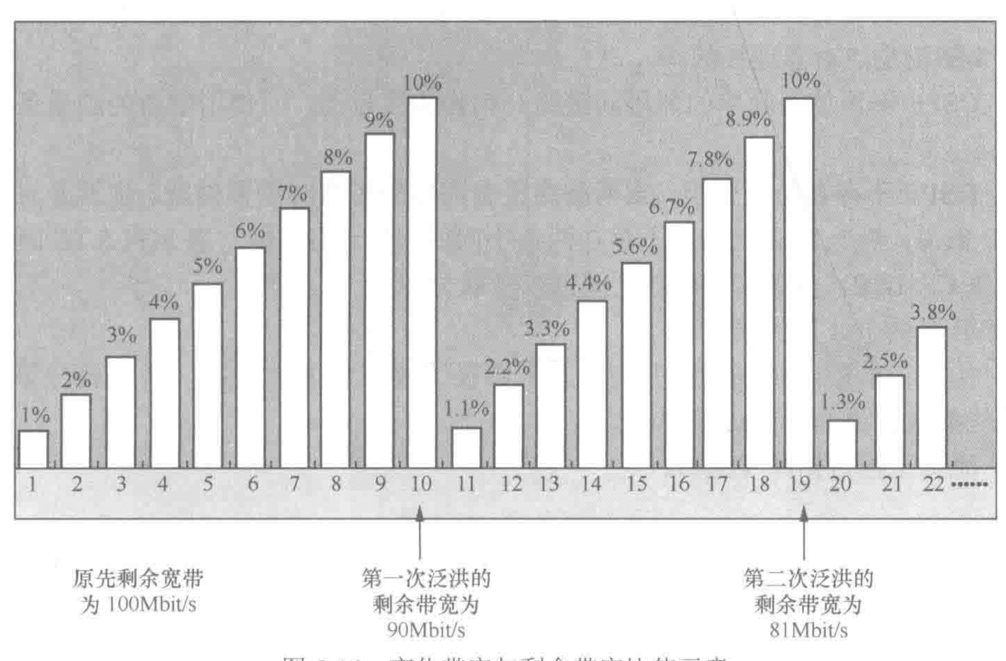


图 5-14 变化带宽与剩余带宽比值示意

建立第 1～9 条时，不进行泛洪；当建立第 10 条时才对第 1～10 条所占用的 10Mbit/s 带宽进行泛洪。此时剩余带宽为 90Mbit/s。当建立第 11～18 条隧道时不进行泛洪，当建立第 19 条时才泛洪。依此类推。

OSPF TE 或 IS-IS TE 泛洪完成后，将形成本区域内统一的流量工程数据库 TEDB。

设备根据 TEDB 中的信息计算 Egress 节点到达区域内其他节点的最合适的路径。MPLS TE 使用该路径建立 CR-LSP。

说明 TEDB 与 IGP 路由协议的 LSDB 是两个完全独立的数据库。两者来源相同，都是 IGP 路由协议泛洪的产物。但内容和功能不同，TEDB 除了具备 LSDB 中所有的内容外，还包含流量工程的信息。LSDB 用于 IGP 最短路径的计算而 TEDB 则用于流量工程 LSP 最优路径的计算。

5.3 CR-LSP 路径计算

MPLS TE 使用 CSPF 算法计算出到达某个 Egress 节点的最优路径。CSPF 的计算过程就是根据 MPLS TE 隧道的要求，先对 TEDB 中的链路进行裁剪，把不满足 TE 属性要求的链路剪掉；然后再采用 SPF 算法，寻找一条到达 Egress 节点、满足 TE 属性要求的最短路径（即一组 LSR 地址）。CSPF 计算的结果是一条满足约束条件、完全明确的路径，通常只在 MPLS TE 隧道的 Ingress 节点进行 CSPF 计算。

1. CSPF 与 SPF

CSPF 是专门用于 MPLS TE 路径计算的算法，它与一般的 SPF 算法相差不大，但又有几点区别。

■ CSPF 只计算到达隧道终点的最短路径，而 SPF 需要计算到达所有节点的最短路径，因为隧道是点对点连接的。

■ CSPF 不再使用简单的邻居间链路开销作为度量值，而使用隧道的约束条件作为度量值。

■ **CSPF 不存在负载分担，当两条路径有同样的权值时需要仲裁**。也就是对于一条 CR-LSP 来说，整个 LSP 上始终不存在两条不同的转发路径。但一条 MPLS TE 隧道可以包含多条 CR-LSP，多条 CR-LSP 之间可以负载分担。

CSPF 有两个计算依据。

■ 待建立 CR-LSP 隧道的带宽、显式路径、建立/保持优先级、亲和属性等约束条件，这些都在隧道的 Ingress 节点配置。

■ 流量工程数据库 TEDB。

说明 如果网络中没有配置 OSPF TE 和 IS-IS TE，就不能形成 TEDB，但仍可以由 IGP 生成 CR-LSP（不是 CSPF 计算出来的）。当然，此时建立的其实不是 CR-LSP，而是普通的 LSP。

2. CSPF 仲裁

上文提到，CSPF 在计算路径的过程中，如果遇到多条权值相同的路径，将根据策略仅选择其中的一条。这个过程称为仲裁（tie-breaking），可用的仲裁策略有。

■ Most-fill：选择已用带宽和最大可预留带宽的比值最大的链路，这样就使流量尽可能选择已使用的链路，使链路带宽资源高效使用。

■ Least-fill：选择已用带宽和最大可预留带宽的比值最小的链路，这样就使流量尽可能选择还未使用的链路，使各条链路的带宽资源均匀使用。

■ Random：随机选取，使每条链路上的 LSP 数量均匀分布，不考虑带宽因素。

在已用带宽和最大可预留带宽的比值相同的情况下，如果存在多格可选链路，则此时不管配置的是 Least-fill 还是 Most-fill 策略，最终选择的是首先发现的链路。

3. CSPF 计算过程

下面以图 5-15 所示的上图为例来说明 CSPF 的计算过程。除了标注具体带宽以及蓝

色（Blue）的链路外，其他链路都为 100Mbit/s。此时需要建立一条目的地址为 LSRE，带宽为 80Mbit/s，且必须经过 LSRH 节点的 MPLS TE 隧道。

说明　本示例中所说的标注为“Blue”的链路其实可以看作是在 CR-LSP 计算中明确要求避开的链路。

因为 50Mbit/s 链路不符合隧道带宽需求，而标注为 Blue 的链路明确被排除在外，所以在链路选择时会裁剪掉这些链路，但又明确要求必须经过 LSRH，所以经过 CSPF 计算裁剪后，得到的链路拓扑如图 5-15 中的下图所示。

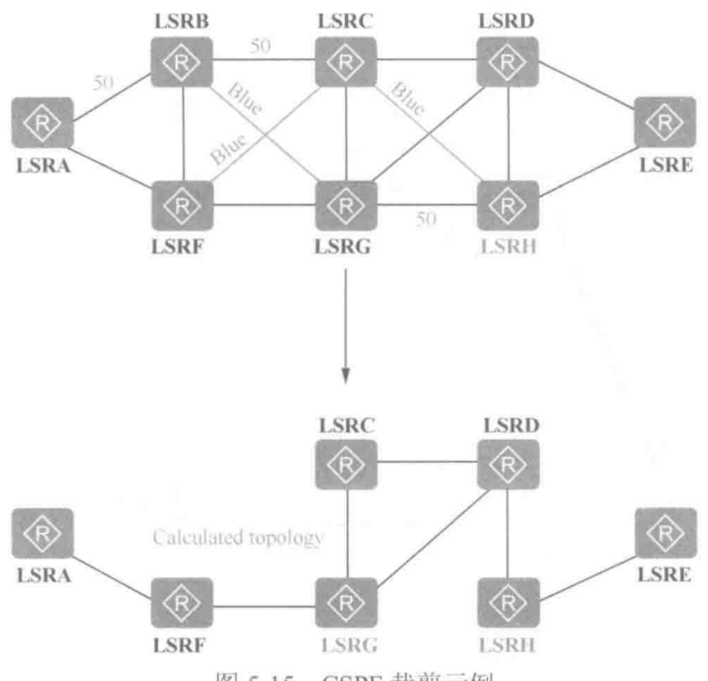


图 5-15　CSPF 裁剪示例

说明　图中的 LSRB 与 LSRF 之间的链路虽然符合要求，但因为 LSRB 上的其他链路均被排除在外了，从 LSRA 发出的报文经 LSRF 到达 LSRB 后没有出接口可选，于是就没有意义了，所以最终 LSRB 也不会包含在 CR-LSP 计算范围之内。

然后再经过 OSPF TE 或者 IS-IS TE 进一步按照普通的 SPF 进行拓扑计算，因为 LSRG、LSRC 和 LSRD 之间存在环路，假设根据 SPF 计算在他们之间找到一条最优路径：LSRG→LSRD，所以最终得到如图 5-16 所示的计算结果。

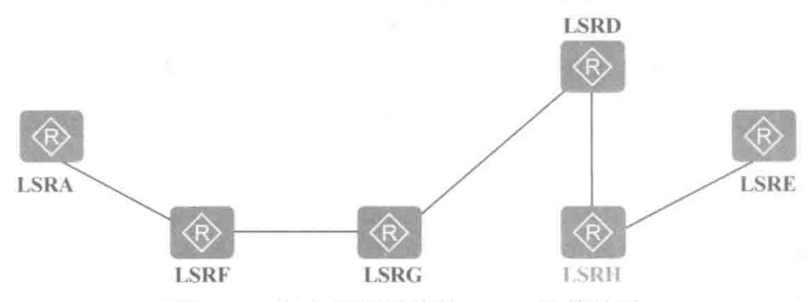


图 5-16　以上示例最终的 CSPF 计算结果

5.4 CR-LSP 路径的建立与切换

使用 CSPF 算法计算出满足约束条件的路径后，MPLS TE 通过标签分发协议 RSVP-TE 沿着计算出的路径建立 CR-LSP，并在路径经过的节点上预留资源。另外，建立好的 CR-LSP 还可能在使用的过程中需要改变路径，这就涉及 CR-LSP 路径的切换。本节分别予以介绍。

5.4.1 CR-LSP 的路径建立原理

CR-LSP 的建立方式可以分为静态建立和动态建立两种。因为静态 CR-LSP 的建立完全依靠企业网络管理员的手工配置，故此处仅介绍使用 RSVP-TE 信令建立动态 CR-LSP 的原理。

动态 CR-LSP 的建立主要分为两个步骤：（1）Ingress 节点向 Egress 节点发送 Path 消息；（2）Egress 节点向 Ingress 节点发送 Resv 消息。中间如果发生错误，则会触发对应的 Err（如 PathErr 或 ResvErr）消息的发送。Path 消息用于创建 RSVP 会话和关联路径状态，接收了 Path 消息的途经节点会建立路径状态块 PSB。Resv 消息携带了资源预留信息，发送时途经的节点会建立资源预留状态块 RSB 和分配标签。

在正式建立 CR-LSP 之前，需要在 RSVP-TE 隧道的入节点完成 CSPF 的路径计算，参见 5.3 节。CSPF 是 MPLS TE 路径选择的核心，与 SPF 类似，只是在计算最优路径时把预留带宽也考虑进来。需要注意的是，在 SPF 中到同一目的地可以有很多跳代价相同的路径，但 CSPF 并不是要计算到一个目的地的所有可能路径，而是对于一个目的地只有一条路径，如果遇到了具有相同开销的等价路径时 CSPF 需要进行一个仲裁，从中选出一条路径作为最终计算的结果。另外，RSVP-TE 不是路由协议，因此 CSPF 路径的计算需要借助于基于状态信息的路由协议（OSPG、ISIS 等）生成的路由信息。

下面以图 5-17 中从 PE1 到 PE2 动态建立 CR-LSP 为例，详细介绍 RSVP-TE 建立 CR-LSP 的流程。

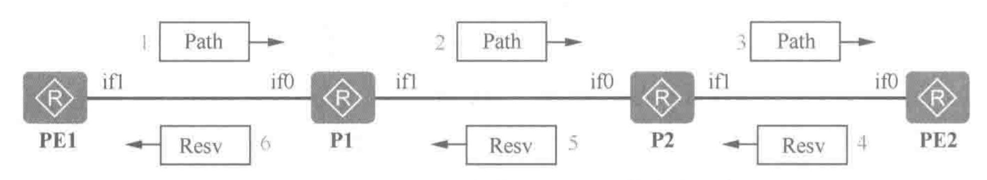


图 5-17 RSVP-TE 建立 CR-LSP 的流程示例

首先，PE1 按照 5.3 节介绍的 CR-LSP 路径计算原理，触发 CSPF 计算从 PE1 到 PE2 的路径，这条路径是指定了沿途每一跳的 IP 地址。然后才正式进行下面的 CR-LSP 路径建立流程。

（1）PE1 将 CSPF 计算出来的 IP 地址列表作为 ERO（显式路径对象）的内容，构造 Path 消息，并根据 Path 消息构造 PSB，然后根据 ERO 将 Path 消息发送给 P1，Path 消息携带的内容见表 5-7（包括多个消息对象）。

表 5-7　**PE1 节点的 Path 消息**

Object	Value
SESSION（会话信息，包括源端和目的端）	Source：PE1-if1；Destination：PE2-if0
RSVP_HOP（发送 Path 消息的出接口）	PE1-if1
EXPLICIT_ROUTE（LSP 路径信息）	P1-if0；P2-if0；PE2-if0
LABEL（标签请求对象）	LABEL_REQUEST

（2）P1 收到 PE1 的 Path 消息后，解析报文，根据 Path 消息构建自己的 PSB。然后 P1 更新 Path 消息，根据 ERO 将 Path 消息发送给 P2，Path 消息携带的内容见表 5-8。P1 的主要操作如下。

- P1 更新 Path 消息的 RSVP_HOP 为 P1 到 P2 的出接口地址。
- P1 更新 Path 消息的 ERO，删除 P1 自己的出、入接口地址及 LSR ID。

表 5-8　**P1 节点的 Path 消息**

Object	Value
SESSION	Source：PE1-if1；Destination：PE2-if0
RSVP_HOP	P1-if1
EXPLICIT_ROUTE	P2-if0；PE2-if0
LABEL	LABEL_REQUEST

（3）P2 收到 P1 的 Path 消息后，与 P1 类似的处理，根据 Path 消息构建自己的 PSB，并更新 Path 消息，删除 P2 自己的出、入接口地址及 LSR ID，根据 ERO 将 Path 消息发送给 PE2，Path 消息携带的内容见表 5-9。

表 5-9　**P2 节点的 Path 消息**

Object	Value
SESSION	Source：PE1-if1；Destination：PE2-if0
RSVP_HOP	P2-if1
EXPLICIT_ROUTE	PE2-if0
LABEL	LABEL_REQUEST

（4）PE2 收到 Path 消息后，从 Session 对象的 Destination 内容获知自己为待建立的 CR-LSP 的 Egress 节点。此时，PE2 分配标签（为入标签）和资源，根据本地 PSB 产生 Resv 消息。Resv 消息被发送给 P2，该消息携带了 PE2 分配给 P2 的标签（为出标签），Resv 消息携带的内容见表 5-10。

表 5-10　**PE2 节点的 Resv 消息**

Object	Value
SESSION	Source：PE2-if0；Destination：PE1-if1
RSVP_HOP	PE2-if0
LABEL	3
RECORD_ROUTE	PE2-if0

说明　各节点在收到 Path 消息并构建自己的 PSB 后，需对资源进行检查，包括带宽是否满足要求、亲和属性是否正确等。如果检查出错，则需要根据错误类型向上游节点发

送 PathErr 消息。

PE2 从 PSB 提取 RSVP_HOP 字段的地址作为 Resv 消息的目的 IP 地址，直接沿着原来 Path 消息发送的逆向路径进行转发，因此 Resv 消息中并不携带 ERO。后面的节点一样。

Resv 消息如果包含 RESV_CONFIRM 对象，接收该消息的节点需要向发送该消息的节点发送 ResvConf 消息来确认资源预留请求。

（5）当 P2 收到 Resv 消息后，记录相关信息到 RSB 中，并分配一个新标签，更新 Resv 消息发给 P1，Resv 消息携带的内容见表 5-11。

表 5-11　　P2 节点的 Resv 消息

Object	Value
SESSION	Source：PE2-if0；Destination：PE1-if1
RSVP_HOP	P2-if0
LABEL	17
RECORD_ROUTE	P2-if0；PE2-if0

（6）当 P1 收到 P2 的 Resv 消息时，同 P2 一样，记录相关信息到 RSB 中，并分配一个新标签，更新 Resv 消息发给 PE1，Resv 消息携带的内容见表 5-12。

表 5-12　　P1 节点的 Resv 消息

Object	Value
SESSION	Source：PE2-if0；Destination：PE1-if1
RSVP_HOP	P1-if0
LABEL	18
RECORD_ROUTE	P1-if0；P2-if0；PE2-if0

（7）PE1 收到 Resv 消息，获得 P1 分配的标签，表明资源预留成功，此时 CR-LSP 建立成功。

5.4.2 CR-LSP 路径切换的 Make-Before-Break 机制

对于一条建立好的 MPLS TE 隧道而言，当链路属性或隧道属性变化导致有了更优的路径时，原隧道会按照新的属性重新建立 CR-LSP，并在建成后将流量切换到新的 CR-LSP 上。但在实际的应用中，很可能出现在新的 CR-LSP 尚未建立完成时就把流量切换到新路径的情况，但由于新、旧 CR-LSP 可能存在部分重合链路，从而出现资源预留竞争。此时如果链路上可用的带宽不足以同时支持新、旧 CR-LSP 带宽预留时，新路径就无法建立，最终使切换后的流量丢失。

MPLS TE 提供了 Make-Before-Break 机制来避免上述问题的发生，因为 Make-Before-Break 可通过 5.1.2 节介绍的 SE（共享显示）风格使得新、旧路径中重合的节点为多条 CR-LSP 采用同一个资源预留，不会出现多条 CR-LSP 带宽资源预留的竞争。这样一来，新路径需要预留的带宽不被重复计算，即采用原路径使用的带宽。即路径重合的地方不额外占用带宽，路径不重合的地方还是额外占用带宽。

如图 5-18 所示，假设所有链路最大可预留带宽为 60Mbit/s。原来已建立了一条由

Router_1到Router_4的CR-LSP，带宽为40Mbit/s，路径是Path1。

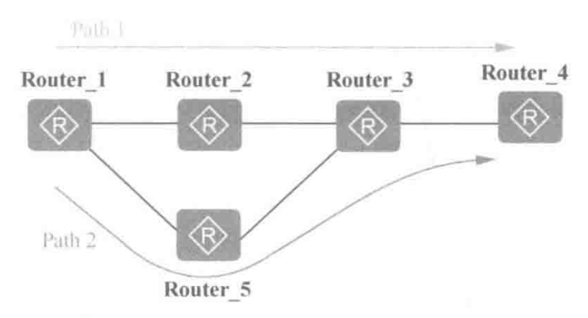


图5-18 Make-Before-Break机制示例

现在希望将路径改为Path2，通过负载较轻的Router_5进行数据转发。此时Router_3→Router_4上剩余的可预留带宽只有20Mbit/s，不足为新路径重新预留40Mbit/s带宽。在这种情况下，可以通过Make-Before-Break机制来解决，新建立的路径Path2在Router_3→Router_4上进行资源预留时采用原路径使用的带宽。新隧道建立成功后，流量转到新路径上后拆除原路径。

与此类似的是增加隧道的带宽，只要共用链路的可预留带宽满足增量要求，新的CR-LSP就可以建立成功。

仍以图5-18为例，假设所有链路最大可预留带宽为60Mbit/s。原来路径Path1的带宽为30Mbit/s。现在希望将路径改为Path2，通过负载较轻的Router_5进行数据转发，并将带宽增大为40Mbit/s。此时Router_3→Router_4上剩余的可预留带宽只有30Mbit/s，不足以单独再为新路径预留40Mbit/s带宽。在这种情况下，也可以通过Make-Before-Break机制来解决，让新建立的Path2在Router_3→Router_4上进行资源预留时采用原路径使用的带宽，并追加增量带宽。这样，新的CR-LSP建立成功后，也会在流量转到新路径上后拆除原路径。

5.5 MPLS TE流量转发

通过上文介绍的MPLS TE信息发布、CR-LSP路径计算和CR-LSP路径建立三大功能，已经可以成功建立一条MPLS TE隧道。但不同于LDP LSP，MPLS TE建立好后仍不能自动将流量引入到隧道中进行转发，需要采用一定的方式将流量引入到MPLS TE隧道中进行转发。这里介绍以下几种将流量引入MPLS TE隧道的方式。

- 静态路由指定：适用于网络拓扑简单或者网络环境稳定的场景。
- 策略路由指定：适用于负载分担和安全监控等场景。
- 隧道策略指定：适用于需要选择TE隧道承载VPN业务的场景。
- 自动路由发布：适用于网络拓扑复杂或者网络环境经常变动的场景。

1. 静态路由指定

将流量引入TE隧道最简单的方法是使用静态路由，配置方法很简单，只需将TE隧道的Tunnel接口设置为静态路由的出接口即可。

2. 策略路由指定

PBR（Policy-Based Routing，策略路由）是一种依据用户制定的策略进行路由选择的机制，可应用于安全、负载分担等场景。在 MPLS 网络中，可使符合过滤条件的 IP 报文通过指定的 CR-LSP 转发，策略路由的配置方法与普通 IP 网络中的策略路由配置方法一样，参见《华为路由器学习指南》。

MPLS TE 的策略路由通过定义一系列匹配的规则和动作，将 **apply** 语句的出接口设置为 TE 隧道的接口。如果报文不匹配策略路由规则，将进行正常 IP 转发；如果报文匹配策略路由规则，则报文直接从指定隧道转发。

3. 隧道策略指定

通常，VPN 流量通过隧道进行转发时，默认采用 LSP 隧道而非 MPLS TE 隧道，此时需要在 VPN 应用隧道策略（Tunnel Policy）将 VPN 流量引入到 MPLS TE 隧道中。可以采用以下两种方式来配置隧道策略（有关隧道策略的详细介绍参见《华为 MPLS VPN 学习指南》）。

■ 按优先级顺序选择（Select-seq）方式：该策略可以改变 VPN 选择的隧道类型，按照配置的隧道类型优先级顺序将 TE 隧道选择为 VPN 的公网隧道。

■ 隧道绑定（Tunnel Binding）方式：该策略可以为 VPN 绑定 TE 隧道以保证 QoS，将某个目的 IP 地址与某条 TE 隧道进行绑定。

4. 自动路由发布

自动路由（Auto Route）是指将 MPLS TE 隧道看作逻辑链路，与物理链路一起参与 IGP 路由的计算，并使用对应的隧道接口作为路由出接口。这时，TE 隧道被看作点到点链路，并且可以设置其 Metric（度量）值。自动路由方式有两种。

■ 转发捷径（IGP Shortcut）：不将这条 CR-LSP 链路发布给邻居节点，这样只有建立了 MPLS TE 隧道的节点可能会利用该隧道转发流量，其他节点因为不会感知这条 MPLS TE 隧道的存在，所以不会使用此隧道。

■ 转发邻接（Forwarding Adjacency）：将这条 CR-LSP 发布给邻居节点，这样其他节点也会感知这条隧道的存在，能够使用此隧道来发流量。

转发邻接是通过在 OSPF 第 10 类 Opaque LSA 的 Remote IP Address 子 TLV 和 IS 可达性 TLV 的 Remote IP Address 子 TLV 中携带邻居 IP 地址来发布该 LSP 的。使用转发邻接时，隧道两端必须在同一区域中。

下面通过图 5-19 所示的例子来理解这两种自动路由方式的区别。

在 Router_7 上建立一条目的地址为 Router_2、路径为 Router_7→Router_6→Router_2 的 TE 隧道，且设置此隧道的 TE Metric 为图中所示的值。现在需要在 Router_5 和 Router_7 上分别查询去往 Router_2 和 Router_1 的路由，则有如下情形。

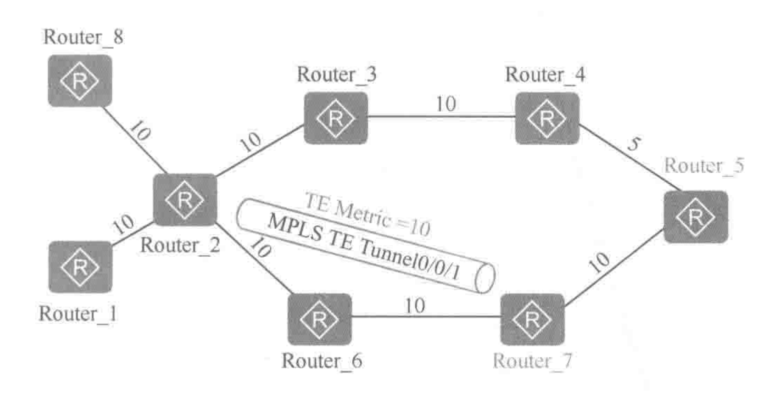
■ 不配置自动路由：去往 Router_2 和 Router_1 的路由的下一跳为 Router_4 和 Router_6。

■ 配置自动路由将流量引入。

- 采用转发捷径方式发布 TE 隧道 Tunnel0/0/1，在 Router_5 和 Router_7 上分别查询去往 Router_2 和 Router_1 的路由，Router_5 的下一跳仍为 Router_4，Router_7 的下一跳变为了 Tunnel0/0/1。可见 Router_5 并不感知隧道的存在，并未利用

Tunnel0/0/1 来进行 IGP 选路，只有 Router_7 自己感知并利用 Tunnel0/0/1 进行 IGP 的选路。

- 采用转发邻接方式发布 TE 隧道 Tunnel0/0/1，在 Router_5 和 Router_7 上分别查询去往 Router_2 和 Router_1 的路由，Router_5 的下一跳变为了 Router_7，Router_7 的下一跳变为了 Tunnel0/0/1。可见 Router_5 和 Router_7 都感知并利用 Tunnel0/0/1 进行 IGP 的选路。



节点	方式	Destination	Nexthop	Cost
Router_5	转发捷径	Router_2	Router_4	25
		Router_1	Router_4	35
Router_7		Router_2	Tunnel0/0/1	10
		Router_1	Tunnel0/0/1	20
Router_5	转发邻接	Router_2	Router_7	20
		Router_1	Router_7	30
Router_7		Router_2	Tunnel0/0/1	10
		Router_1	Tunnel0/0/1	20

图 5-19　转发捷径与转发邻接自动路由方式示例

5.6　静态 MPLS TE 隧道配置与管理

上文介绍的 RSVP-TE、OSPF TE、IS-IS TE、CSPF 和 CR-LSP 路径计算等内容都是针对动态 MPLS TE 隧道（也即动态 CR-LSP）建立而言的，但 MPLS TE 隧道还可手工静态配置，就像 LSP 可以由 LDP 协议自动建立，还可以手动配置一样。在静态 DS-TE 隧道中，所有参数都是手工配置的，不需要以上这些协议参与。

配置静态 MPLS TE 隧道可以实现静态 CR-LSP 的建立。配置过程比较简单，手工分配标签，不使用信令协议（即不需要使用 RSVP-TE），不需要通过 OSPF TE 或 IS-IS TE 协议交互控制报文，也不需要通过 CSPF 计算路由，因此消耗资源比较小。

配置静态 MPLS TE 隧道的任务如，其中配置链路的带宽为可选配置任务。

（1）使能 MPLS TE

需要在骨干网各节点全局和公网接口上使能 MPLS TE，这是建立 MPLS TE 隧道的

基础和前提。

（2）配置 MPLS TE 隧道接口

要创建 MPLS TE 隧道，必须先创建一个 Tunnel 接口，然后在 Tunnel 接口下完成隧道的其他属性配置。隧道的接口主要负责隧道的建立、管理和指导报文转发。

（3）（可选）配置链路的带宽

可以为建立的 MPLS TE 隧道配置一定的带宽约束条件，包括链路最大可预留带宽，以及为链路上各条隧道上传输的两类业务分配的带宽（BC0 和 BC1）。

（4）在骨干网各节点上手动创建静态 CR-LSP，在一路通信中必须创建双向的静态 CR-LSP。

当静态 CR-LSP 绑定到隧道接口后，通过该隧道接口转发的流量直接进入该 CR-LSP 中转发，不需要依赖 IP 路由即可生效。如何使不同用户流量通过指定的隧道接口转发，将在 5.8 节介绍。

在配置静态 MPLS TE 隧道之前，需要完成以下任务。

- 配置各 LSR 的 LSR-ID。
- 在各 LSR 节点上全局和公网接口下使能 MPLS。

5.6.1 使能 MPLS TE

使能 MPLS TE 功能需要在 MPLS TE 隧道中各节点上进行配置，具体配置方法见表 5-13。但在使能 MPLS TE 功能之前必须先使能 MPLS 功能。

表 5-13 使能 MPLS TE 的配置步骤

步骤	命令	说明
1	**system-view** 例如：<Huawei> **system-view**	进入系统视图
2	**mpls** 例如：[Huawei] **mpls**	使能全局 MPLS 功能，进入 MPLS 视图
3	**mpls te** 例如：[Huawei-mpls] **mpls te**	全局使能本节点的全局 MPLS TE 功能，必须先全局使能 MPLS TE 功能，才能在接口下使能 MPLS TE 功能。 缺省情况下，未使能全局 MPLS TE 功能，可用 **undo mpls te** 命令去使能 MPLS TE 功能，所有接口的 MPLS TE 也同时被去使能，所有的 CR-LSP 将被删除
4	**quit** 例如：[Huawei-mpls] **quit**	返回系统视图
5	**interface** *interface-type interface-number* 例如：[Huawei] **interface** gigabitethernet 1/0/0	进入 MPLS/IP 公网接口的接口视图
6	**mpls** 例如：[Huawei-GigabitEthernet1/0/0] **mpls**	在以上公网接口上使能 MPLS
7	**mpls te** 例如：[Huawei-GigabitEthernet1/0/0] **mpls te**	在以上公网接口上使能 MPLS TE。 缺省情况下，未使能接口的 MPLS TE 功能，可用 **undo mpls te** 命令去使能以上接口的 MPLS TE 功能，当前接口上的所有 CR-LSP 也将变为 Down

5.6.2 配置 MPLS TE 隧道接口

要创建 MPLS TE 隧道，必须先创建一个 Tunnel 接口，然后在 Tunnel 接口下完成隧道的其他属性配置，具体配置方法见表 5-14，**但仅需在 Ingress 节点上配置**。隧道的接口主要负责隧道的建立、管理和指导报文转发。

说明 由于 MPLS TE 隧道转发的是 MPLS 报文，因此，在隧道接口下配置 IP 报文转发相关的命令是无效的。例如源地址合法性检查 **ip verify source-address** 命令、单播逆向路径转发 URPF 检查 **urpf** 命令。

表 5-14　**MPLS TE 隧道接口的配置步骤**

步骤	命令	说明
1	**system-view** 例如：<Huawei> **system-view**	进入系统视图。
2	**interface tunnel** *interface-number* 例如：[Huawei] **interface tunnel** 0/0/1	创建 Tunnel 接口并进入 Tunnel 接口视图。参数 *interface-number* 用来指定 Tunnel 接口的编号，格式为“槽位号/卡号/端口号”，槽位号、卡号均为整数形式，取值与设备有关；端口号为整数形式。 缺省情况下，系统未创建 Tunnel 接口，可用 **undo interface tunnel** *interface-number* 命令删除指定的 Tunnel 接口
3	**ip address** *ip-address* { *mask* | *mask-length* } [**sub**] 例如：[Huawei-Tunnel0/0/1] **ip address** 10.1.1.1 24	（二选一）配置 Tunnel 接口的 IP 地址
	ip address unnumbered interface *interface-type interface-number* 例如：[Huawei-Tunnel0/0/1] **ip address unnumbered interface** loopback 0	（二选一）配置隧道接口借用其他接口的 IP 地址。参数 *interface-type interface-number* 指定被借用的接口。 **【说明】**如果 Tunnel 接口不配置 IP 地址，不影响 TE 隧道的成功建立。但是如果需要实现流量转发，则必须为 Tunnel 接口配置 IP 地址。通常的做法是在 Ingress 节点创建一个 Loopback 接口并配置与 LSR ID 相同的 32 位 IP 地址，然后 Tunnel 接口借用该 Loopback 接口的 IP 地址。 缺省情况下，接口不借用其他接口的 IP 地址，可用 **undo ip address unnumbered** 命令取消接口借用其他接口的 IP 地址
4	**tunnel-protocol mpls te** 例如：[Huawei-Tunnel0/0/1] **tunnel-protocol mpls te**	配置隧道协议为 MPLS TE。 缺省情况下，Tunnel 接口的隧道协议为 **none**，即不进行任何协议封装，可用 **undo tunnel-protocol** 命令恢复缺省配置
5	**destination** *dest-ip-address* 例如：[Huawei-Tunnel0/0/1] **destination** 4.4.4.9	配置隧道的目的地址。参数 *dest-ip-address* 用来指定隧道的目的 IP 地址。一般将隧道的目的 IP 地址配置为出节点的 LSR ID。 **【说明】**由于不同类型的隧道对于目的地址要求不同，当隧道协议从其他类型改变为 MPLS TE 时，原

（续表）

步骤	命令	说明
5	**destination** *dest-ip-address* 例如：[Huawei-Tunnel0/0/1] **destination** 4.4.4.9	先配置的隧道目的 IP 地址将被自动删除，需要重新配置。 缺省情况下，没有配置 Tunnel 接口的目的地址，可用 **undo destination** 命令删除隧道的目的地址
6	**mpls te tunnel-id** *tunnel-id* 例如：[Huawei-Tunnel0/0/1] **mpls te tunnel-id** 100	配置隧道 ID。参数 *tunnel-id* 用来指定隧道 ID，整数形式，取值范围是 1～4096。 隧道 ID 用来唯一标识一条 MPLS TE 隧道，以便对隧道进行规划和管理，Tunnel ID 为隧道的必配项，如果不配置，隧道将无法建立成功
7	**mpls te signal-protocol cr-static** 例如：[Huawei-Tunnel0/0/1] **mpls te signal-protocol cr-static**	配置隧道使用静态 CR-LSP。 缺省情况下，MPLS TE 建立隧道使用 RSVP-TE 信令协议，可用 **undo mpls te signal-protocol cr-static** 命令恢复缺省配置。
8	**mpls te signalled tunnel-name** *tunnel-name* 例如：[Huawei-Tunnel0/0/1] **mpls te signalled tunnel-name** LSRAtoLSRC	（可选）配置 TE 隧道的名称。参数 *tunnel-name* 用来指定 TE 隧道的名称，字符串形式，长度范围是 1～63，不支持空格和“/”，**区分大小写**。首字符必须为“_”或者字母，不能是数字。 缺省情况下，TE 隧道的名称用隧道接口的名称来标识，如 Tunnel0/0/1，可用 **undo mpls te signalled tunnel-name** 命令恢复缺省配置
9	**mpls te commit** 例如：[Huawei-Tunnel0/0/1] **mpls te commit**	提交隧道当前配置。 **【注意】**在执行本命令之前，对 MPLS TE 隧道的配置不会生效。如果未配置 destination 或隧道 ID，则提交失败，需要补全后再重新提交。当 MPLS TE 的参数发生改变，需要使用该命令使之生效。每次更改 Tunnel 接口上的 MPLS TE 参数后，都需要使用本命令提交配置

5.6.3 配置链路的带宽

配置链路的带宽是一可选配置任务，仅当所建立的 MPLS TE 隧道有带宽约束要求时才需要配置。这时需要在 MPLS TE 隧道途经的各节点上配置链路的带宽与其进行协商，使得这条具有带宽约束的 CR-LSP 可以被建立，从而合理利用网络资源。

链路带宽的具体配置步骤见表 5-15，需要在 MPLS TE 隧道的各节点上配置。P 节点的双向链路都要配置，Egress 节点上无需配置。

表 5-15 链路的带宽的配置步骤

步骤	命令	说明
1	**system-view** 例如：<Huawei> **system-view**	进入系统视图。
2	**interface***nterface-type interface-number* 例如：[Huawei] **interface** gigabitethernet 1/0/0	进入使能了 MPLS TE 的接口的视图。

（续表）

步骤	命令	说明
3	**mpls te bandwidth max-reservable-bandwidth** *bw-value* 例如：[Huawei-GigabitEthernet1/0/0] **mpls te bandwidth max-reservable-bandwidth** 10000	配置链路最大可预留带宽，**只需在有带宽要求的 TE 隧道所经链路的各个出方向接口上配置即可**。参数 *bw-value* 用来指定链路的最大可预留带宽，整数形式，取值范围是 0～4000000000，单位是 kbit/s。缺省值是 0。 缺省情况下，链路的最大可预留带宽为 0bit/s，可用 **undo mpls te bandwidth max-reservable-bandwidth** 命令用来恢复系统缺省配置。当 MPLS TE 隧道入节点发起建立具有带宽约束的 CR-LSP 时，如果不配置链路的最大可预留带宽，那么要求的 CR-LSP 带宽就会大于链路最大可预留带宽，CR-LSP 也就无法建立成功
4	**mpls te bandwidth** { **bc0** *bc0-bw-value* | **bc1** *bc1-bw-value* }* 例如：[Huawei-GigabitEthernet1/0/0] **mpls te bandwidth bc0** 1000	配置链路的 BC（Bandwidth Constraint，带宽约束）带宽，**只需在有带宽要求的 TE 隧道所经链路的出方向接口上配置即可**。BC 是指一条 MPLS TE 隧道中一类 CT（Class Type）的总带宽，这方面涉及 DS-TE 隧道，具体将在本书第 9 章介绍。命令中的参数说明如下。 • **bc0** *bc0-bw-value*：可多选参数，设置 BC0 的带宽，整数形式，取值范围是 1～4000000000，单位是 Kbit/s。缺省值是 1。 • **bc1** *bc1-bw-value*：可多选参数，设置 BC1 的带宽，整数形式，取值范围是 1～4000000000，单位是 Kbit/s。缺省值是 1。 【说明】如果要修改 BC 带宽值，可重新配置该命令，最后一次的配置会覆盖之前的配置，但不允许将 BC 带宽值修改成小于已经分配给 CT 的带宽值。例如，BC0 已有 10Mbit/s 的带宽被分配，则 BC0 的带宽只能修改成大于等于 10Mbit/s。 如需要进行精确的带宽控制，则需要配置链路上的 BCi 带宽值不小于经过该链路的 DS-TE 隧道上所有 CTi（$0 \leqslant i \leqslant 7$）的带宽总和的 125%。 缺省情况下，没有配置链路的 BC 带宽，可用 **undo mpls te bandwidth** 命令恢复系统缺省配置

5.6.4 配置静态 CR-LSP

配置静态 MPLS TE 隧道时，需要分别在 MPLS TE 隧道入节点、中间节点和出节点处手工配置静态 CR-LSP。当没有中间节点时，可以不必配置中间节点的静态 CR-LSP。

说明 在各节点上需要配置的 CR-LSP 标签与 LDP LSP 标签一样，Ingress 节点上只需配置出标签，Transit 节点上需要同时配置入标签和出标签，Egress 节点上只需配置入标签。但上游节点的出标签要与下游节点的入标签一致。

1．在 Ingress 节点上的配置

在系统视图下执行 **static-cr-lsp ingress** { **tunnel-interface tunnel** *interface-number* | *tunnel-name* } **destination** *destination-address* { **nexthop** *next-hop-address* | **outgoing-interface** *interface-type interface-number* } * **out-label** *out-label* 命令，配置入节点的静态 CR-LSP。命令中的参数说明如下。

■ **Tunnel** *interface-number*：二选一参数，指定静态 CR-LSP 的隧道接口的编号，格式为“槽位号/卡号/端口号”，槽位号、卡号均为整数形式，取值与设备有关，端口号为整数形式。

■ *tunnel-name*：二选一参数，指定静态 CR-LSP 隧道的名称，字符串形式，**区分大小写**，不支持空格和缩写，长度范围是 1～19，必须与命令 **interface tunnel** *interface-number* 创建的隧道接口名称一致，**且区分大小写**，不支持空格。假设使用 **interface Tunnel** 0/0/1 命令为静态 CR-LSP 创建了一个 Tunnel 接口，则入节点中的该参数应该写作“Tunnel0/0/1”，否则隧道将不能正确建立。中间节点和出节点无此限制。

■ *destination-address*：指定静态 CR-LSP 的目的地址，通常为 Egress 节点的 Loopback 接口 IP 地址。

■ **nexthop** *next-hop-address*：可多选参数，指定静态 CR-LSP 下一跳地址。下一跳或出接口由入节点到出节点的路由决定。如果 LSP 出接口为以太网类型，必须配置 **nexthop** *next-hop-address* 参数以保证 LSP 的正常转发。

注意 如果在配置静态 CR-LSP 时指定了下一跳，则在配置 IP 静态路由时也必须指定下一跳，否则不能建立静态 CR-LSP。

■ **outgoing-interface** *interface-type interface-number*：可多选参数，指定出接口类型和编号，只有点到点链路才能选择配置出接口。

■ out-label *out-label*：指定出标签的值，整数形式，取值范围是 16～1048575。上游节点的出标签也是下游节点的入标签，这点与 LDP LSP 是一样的。

缺省情况下，没有在入口节点配置静态 CR-LSP，可用 **undo static-cr-lsp ingress** { **tunnel-interface tunnel** *interface-number* | *tunnel-name* }命令在入口节点删除配置的静态 CR-LSP。但如果要对除 Tunnel 接口外的其他参数进行修改，可直接重新执行本命令进行配置，不用删除原来的 CR-LSP。

例如要在入节点配置静态 CR-LSP，隧道名字为 Tunnel0/0/1，目的 IP 地址为 10.1.3.1，下一跳 IP 地址为 10.1.1.2，出标签为 237。

```
<Huawei> system-view
[Huawei] static-cr-lsp ingress Tunnel0/0/1 destination 10.1.3.1 nexthop 10.1.1.2 out-label 237
```

2．在 Transit 节点上的配置

在系统视图下执行 **static-cr-lsp transit** *lsp-name* [**incoming-interface** *interface-type interface-number*] **in-label** *in-label* { **nexthop** *next-hop-address* | **outgoing-interface** *interface-type interface-number* } * **out-label** *out-label* [**description** *description*]命令，配置中间节点的静态 CR-LSP。命令中的参数说明如下。

■ *lsp-name*：指定 CR-LSP 隧道的名字，符串形式，区分大小写，不支持空格，长

度范围是 1～19。当输入的字符串两端使用双引号时，可在字符串中输入空格，名称取值没有限制，但不能与该节点上已存在的名称相同。为了清晰，可以使用此静态 CR-LSP 的 MPLS TE 隧道的接口名称，如 Tunnel0/0/1。

■ **incoming-interface** *interface-type interface-number*：可选参数，指定 CR-LSP 的入接口。

■ **in-label** *in-label*：指定入标签的值，整数形式，取值范围是 16～1023。

■ **nexthop** *next-hop-address*：可多选参数，指定下一跳 IP 地址。如果 LSP 出接口为以太网类型，必须配置 **nexthop** *next-hop-address* 参数以保证 LSP 的正常转发。

■ **outgoing-interface** *interface-type interface-number*：可多选参数，指定出接口名称。

■ **out-label** *out-label*：指定出标签的值，整数形式，取值范围是 16～1048575。

注意 因为是出标签与入标签的取值范围不一样，为了确保上游节点的出标签与下游节点的入标签保持一致，需要在配置出标签时，必须不能超过入标签的取值范围 16～1023。

■ **description** *description*：可选参数，对所创建的 CR-LSP 进行描述。

缺省情况下，没有在转发节点配置静态 CR-LSP，可用 **undo static-cr-lsp transit** *lsp-name* 命令在转发节点删除指定的静态 CR-LSP。但如果要对除 CR-LSP 隧道名称外的其他参数进行修改，可直接重新执行本命令进行配置，不用删除原来的 CR-LSP。

例如在中间节点配置静态 CR-LSP，名字为 tunnel39，入接口为 GE1/0/0，入口标签为 123，出接口为 GE2/0/0，出口标签为 253。

```
<Huawei> system-view
[Huawei] static-cr-lsp transit tunnel39 incoming-interface gigabitethernet 1/0/0 in-label 123 outgoing-interface gigabitethernet 2/0/0 out-label 253
```

3. 在 Egress 节点上的配置

在系统视图下执行 **static-cr-lsp egress** *lsp-name* [**incoming-interface** *interface-type interface-number*] **in-label** *in-label* [**lsrid** *ingress-lsr-id* **tunnel-id** *tunnel-id*]，配置出节点的静态 CR-LSP。

■ *lsp-name*：指定 CR-LSP 隧道的名称，取值名称也没有限制，但不能与该节点上已存在的名称相同。为了清晰，可以使用此静态 CR-LSP 的 MPLS TE 隧道的接口名称，如 Tunnel0/0/1。

■ **incoming-interface** *interface-type interface-number*：可选参数，指定入接口。

■ **in-label** *in-label*：指定入标签的值，整数形式，取值范围是 16～1023。

■ **lsrid** *ingress-lsr-id*：可选参数，指定入节点的 LSR ID。

■ **tunnel-id** *tunnel-id*：可选参数，指定隧道标识，整数形式，取值范围是 1～65535。

缺省情况下，没有在出口节点配置静态 CR-LSP，可用 **undo static-cr-lsp egress** *lsp-name* 命令在出口节点删除配置的静态 CR-LSP。同样，如果要对除 CR-LSP 隧道名称外的其他参数进行修改，可直接重新执行本命令进行配置，不用删除原来的 CR-LSP。

如在出节点配置静态 CR-LSP，名字为 tunnel34，入接口是 GE1/0/0，入口标签是 233。

```
<Huawei> system-view
[Huawei] static-cr-lsp egress tunnel34 incoming-interface gigabitethernet 1/0/0 in-label 233
```

5.6.5 静态 CR-LSP 配置管理

已经完成静态 MPLS TE 隧道的所有配置后，可用以下 **display** 命令查看相关配置，验证配置结果。

■ **display mpls static-cr-lsp** [*lsp-name*] [{ **include** | **exclude** } *ip-address mask-length*] [**verbose**]：查看指定或所有静态 CR-LSP 信息。

■ **display mpls te tunnel** [**destination** *ip-address*] [**lsp-id** *ingress-lsr-id session-id local-lsp-id*] [**lsr-role** { **all** | **egress** | **ingress** | **remote** | **transit** }] [**name** *tunnel-name*] [{ **incoming-interface** | **interface** | **outgoing-interface** } *interface-type interface-number*] [**te-class0** | **te-class1** | **te-class2** | **te-class3** | **te-class4** | **te-class5** | **te-class6** | **te-class7**] [**verbose**]：查看指定或所有 MPLS TE 隧道信息。

■ **display mpls te tunnel statistics** 或者 **display mpls lsp statistics**：查看 MPLS TE 隧道或 MPLS LSP 统计信息。

■ **display mpls te tunnel-interface** [**tunnel** *interface-number*]：在静态 CR-LSP 入节点查看隧道接口信息。

5.6.6 静态 MPLS TE 隧道配置示例

如图 5-20 所示，要求分别建立一条 LSRA 到 LSRC 和一条从 LSRC 到 LSRA 的静态 MPLS TE 隧道。

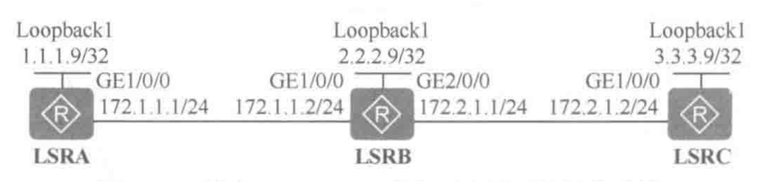


图 5-20 静态 MPLS TE 隧道配置示例的拓扑结构

1. 基本配置思路分析

MPLS TE 的配置是在 MPLS 基础上进行的，所以也需要先配置好 MPLS/IP 骨干网，包括 MPLS/IP 骨干网上的三层路由、全局和公网接口上使能 MPLS，但不要使能 LDP，而是要在全局和公网接口上使能 MPLS TE 功能，然后再配置静态 CR-LSP。

本示例的基本配置思路如下。

（1）配置各节点接口（包括作为 MPLS LSR-ID 的 Loopback 接口）的 IP 地址，并使用 OSPF 协议实现各节点之间公网路由可达。

（2）配置 LSR ID，并全局使能各节点以及公网接口的 MPLS、MPLS TE 功能。

（3）在入节点创建隧道接口，指定使用静态 CR-LSP 建立 MPLS TE 隧道。

（4）在各节点上配置与隧道相关联的静态 CR-LSP，在入节点上配置下一跳地址和出标签，在中间节点配置入接口、下一跳地址和出标签，在出节点上配置入标签和入接口。

注意 要实现双向通信，必须建立两条相反方向的 MPLS TE 隧道，它们的入节点分别为 LSRA、LSRC。每个节点的出标签值要与其下游节点的入标签值一致。

2. 具体配置步骤

（1）配置各设备的各接口的 IP 地址及 OSPF 路由协议。

LSRA 上的配置。

```
<Huawei> system-view
[Huawei] sysname LSRA
[LSRA] interface gigabitethernet 1/0/0
[LSRA-GigabitEthernet1/0/0] ip address 172.1.1.1 255.255.255.0
[LSRA-GigabitEthernet1/0/0] quit
[LSRA] interface loopback 1
[LSRA-LoopBack1] ip address 1.1.1.9 255.255.255.255
[LSRA-LoopBack1] quit
[LSRA] ospf 1
[LSRA-ospf-1] area 0
[LSRA-ospf-1-area-0.0.0.0] network 1.1.1.9 0.0.0.0
[LSRA-ospf-1-area-0.0.0.0] network 172.1.1.0 0.0.0.255
[LSRA-ospf-1-area-0.0.0.0] quit
[LSRA-ospf-1] quit
```

LSRB 上的配置。

```
<Huawei> system-view
[Huawei] sysname LSRB
[LSRB] interface gigabitethernet 1/0/0
[LSRB-GigabitEthernet1/0/0] ip address 172.1.1.2 255.255.255.0
[LSRB-GigabitEthernet1/0/0] quit
[LSRB] interface gigabitethernet 2/0/0
[LSRB-GigabitEthernet2/0/0] ip address 172.2.1.1 255.255.255.0
[LSRB-GigabitEthernet2/0/0] quit
[LSRB] interface loopback 1
[LSRB-LoopBack1] ip address 2.2.2.9 255.255.255.255
[LSRB-LoopBack1] quit
[LSRB] ospf 1
[LSRB-ospf-1] area 0
[LSRB-ospf-1-area-0.0.0.0] network 2.2.2.9 0.0.0.0
[LSRB-ospf-1-area-0.0.0.0] network 172.1.1.0 0.0.0.255
[LSRB-ospf-1-area-0.0.0.0] network 172.2.1.0 0.0.0.255
[LSRB-ospf-1-area-0.0.0.0] quit
[LSRB-ospf-1] quit
```

LSRC 上的配置。

```
<Huawei> system-view
[Huawei] sysname LSRC
[LSRC] interface gigabitethernet 1/0/0
[LSRC-GigabitEthernet1/0/0] ip address 172.2.1.2 255.255.255.0
[LSRC-GigabitEthernet1/0/0] quit
[LSRC] interface loopback 1
[LSRC-LoopBack1] ip address 3.3.3.9 255.255.255.255
[LSRC-LoopBack1] quit
[LSRC] ospf 1
[LSRC-ospf-1] area 0
[LSRC-ospf-1-area-0.0.0.0] network 3.3.3.9 0.0.0.0
[LSRC-ospf-1-area-0.0.0.0] network 172.2.1.0 0.0.0.255
[LSRC-ospf-1-area-0.0.0.0] quit
[LSRC-ospf-1] quit
```

以上配置完成后，LSRA、LSRB、LSRC 之间应能建立 OSPF 邻居关系，执行 **display**

ospf peer 命令可以看到邻居状态为 Full。执行 **display ip routing-table** 命令可以看到 LSR 之间学习到对方的 Loopback1 路由。

（2）配置 MPLS LSR-ID，在全局和公网接口上使能 MPLS 和 MPLS TE 能力。

LSRA 上的配置。

```
[LSRA] mpls lsr-id 1.1.1.9
[LSRA] mpls
[LSRA-mpls] mpls te
[LSRA-mpls] quit
[LSRA] interface gigabitethernet 1/0/0
[LSRA-GigabitEthernet1/0/0] mpls
[LSRA-GigabitEthernet1/0/0] mpls te
[LSRA-GigabitEthernet1/0/0] quit
```

LSRB 上的配置。

```
[LSRB] mpls lsr-id 2.2.2.9
[LSRB] mpls
[LSRB-mpls] mpls te
[LSRB-mpls] quit
[LSRB] interface gigabitethernet 1/0/0
[LSRB-GigabitEthernet1/0/0] mpls
[LSRB-GigabitEthernet1/0/0] mpls te
[LSRB-GigabitEthernet1/0/0] quit
[LSRB] interface gigabitethernet 2/0/0
[LSRB-GigabitEthernet2/0/0] mpls
[LSRB-GigabitEthernet2/0/0] mpls te
[LSRB-GigabitEthernet2/0/0] quit
```

LSRC 上的配置。

```
[LSRC] mpls lsr-id 3.3.3.9
[LSRC] mpls
[LSRC-mpls] mpls te
[LSRC-mpls] quit
[LSRC] interface gigabitethernet 1/0/0
[LSRC-GigabitEthernet1/0/0] mpls
[LSRC-GigabitEthernet1/0/0] mpls te
[LSRC-GigabitEthernet1/0/0] quit
```

（3）配置 MPLS TE 隧道接口，配置隧道接口的 IP 地址、目的 IP 地址，隧道 ID，并指定采用静态 CR-LSP。

在 LSRA 上配置 MPLS TE 隧道接口，用于建立由 LSRA 到 LSRC 的 MPLS TE 隧道。

```
[LSRA] interface tunnel 0/0/1
[LSRA-Tunnel0/0/1] ip address unnumbered interface loopback 1   #---借用 Loopback1 接口的 IP 地址
[LSRA-Tunnel0/0/1] tunnel-protocol mpls te
[LSRA-Tunnel0/0/1] destination 3.3.3.9
[LSRA-Tunnel0/0/1] mpls te tunnel-id 100   #---配置 MPLS TE 隧道 ID
[LSRA-Tunnel0/0/1] mpls te signal-protocol cr-static   #---配置采用静态 CR-LSP
[LSRA-Tunnel0/0/1] mpls te commit   #---提交以上 MPLS TE 配置，使配置生效
[LSRA-Tunnel0/0/1] quit
```

在 LSRC 上配置 MPLS TE 隧道，用于建立由 LSRC 到 LSRA 的 MPLS TE 隧道。

```
[LSRC] interface tunnel 0/0/1
[LSRC-Tunnel0/0/1] ip address unnumbered interface loopback 1
[LSRC-Tunnel0/0/1] tunnel-protocol mpls te
[LSRC-Tunnel0/0/1] destination 1.1.1.9
[LSRC-Tunnel0/0/1] mpls te tunnel-id 200
```

```
[LSRC-Tunnel0/0/1] mpls te signal-protocol cr-static
[LSRC-Tunnel0/0/1] mpls te commit
[LSRC-Tunnel0/0/1] quit
```

（4）在各节点上创建双向静态 CR-LSP。

■ 创建 LSRA 至 LSRC 的静态 CR-LSP。

配置 LSRA 为静态 CR-LSP 的入节点，指定 Tunnel 接口，目的 IP 地址（出节点 LSRC 的 Loopback 接口 IP 地址）、到达 LSRB 的下一跳 IP 地址，分配出标签（20）。

```
[LSRA] static-cr-lsp ingress tunnel-interface Tunnel 0/0/1 destination 3.3.3.9 nexthop 172.1.1.2 out-label 20
```

配置 LSRB 为静态 CR-LSP 的中间节点，指定出接口、分配入标签（20，与 LSRA 的出标签一致）和出标签（30）。

```
[LSRB] static-cr-lsp transit LSRA2LSRC incoming-interface gigabitethernet 1/0/0 in-label 20 nexthop 172.2.1.2 out-label 30
```

配置 LSRC 为静态 CR-LSP 的出节点，指定入接口，分配出标签（30，与 LASRB 上分配的入标签一致）。

```
[LSRC] static-cr-lsp egress LSRA2LSRC incoming-interface gigabitethernet 1/0/0 in-label 30
```

■ 创建 LSRC 至 LSRA 的静态 CR-LSP

配置 LSRC 为静态 CR-LSP 的入节点，指定 Tunnel 接口，目的 IP 地址（出节点 LSRA 的 Loopback 接口 IP 地址）、到达 LSRB 的下一跳 IP 地址，分配出标签（120）。

```
[LSRC] static-cr-lsp ingress tunnel-interface Tunnel0/0/1 destination 1.1.1.9 nexthop 172.2.1.1 out-label 120
```

配置 LSRB 为静态 CR-LSP 的中间节点，指定出接口、分配入标签（120，与 LSRC 的出标签一致）和出标签（130）。

```
[LSRB] static-cr-lsp transit LSRC2LSRA incoming-interface gigabitethernet 2/0/0 in-label 120 nexthop 172.1.1.1 out-label 130
```

配置 LSRA 为静态 CR-LSP 的出节点，指定入接口，分配出标签（130，与 LASRB 上分配的入标签一致）。

```
[LSRA] static-cr-lsp egress LSRC2LSRA incoming-interface gigabitethernet 1/0/0 in-label 130
```

3. 配置结果验证

以上配置全部完成后，在 LSRA 上执行 **display interface tunnel** 命令，可以看到 Tunnel 接口的状态为 Up。以下是在 LSRA 上执行该命令的输出示例。

```
[LSRA] display interface tunnel 0/0/1
Tunnel0/0/1 current state : Up
Line protocol current state : Up
...
```

在各节点上执行 **display mpls te tunnel** 命令，可以看到 MPLS TE 隧道的建立情况。以下是在 LSRA 上执行该命令的输出示例。

```
[LSRA] display mpls te tunnel

-------------------------------------------------------------------------------
Ingress LsrId    Destination      LSPID   In/Out Label    R   Tunnel-name
-------------------------------------------------------------------------------
1.1.1.9          3.3.3.9          1       --/20           I   Tunnel0/0/1
-                -                -       130/--          E   LSRC2LSRA
```

在各节点上执行 **display mpls lsp** 或 **display mpls static-cr-lsp** 命令，可以看到静态 CR-LSP 的建立情况。以下是在 LSRA 上执行这两条命令的输出示例。

```
[LSRA] display mpls lsp
-------------------------------------------------------------------------------
```

```
                    LSP Information: STATIC CRLSP
------------------------------------------------------------------------------
FEC                In/Out Label   In/Out IF                 Vrf Name
3.3.3.9/32         NULL/20        -/GE1/0/0
-/-                130/NULL       GE1/0/0/-

[LSRA] display mpls static-cr-lsp
TOTAL          : 2    STATIC CRLSP(S)
Up             : 2    STATIC CRLSP(S)
DOWN           : 0    STATIC CRLSP(S)
Name           FEC            I/O Label   I/O If              Status
Tunnel0/0/1    3.3.3.9/32     NULL/20     -/GE1/0/0           Up
LSRC2LSRA      -/-            130/NULL    GE1/0/0/-           Up
```

通过以上命令的执行，可以验证以上配置是正确的，双向 MPLS TE 隧道（各仅包括一条 CR-LSP）建立成功了。

5.7 动态 MPLS TE 隧道配置与管理

动态 MPLS TE 隧道的建立需要使用 RSVP-TE 信令协议和 OSPF TE 或 IS-IS TE，还要通过 CSPF 进行路径计算和建立。动态 MPLS TE 隧道可以根据网络变化动态改变，在规模较大的组网中，可以避免逐跳配置的麻烦。配置动态 MPLS TE 隧道是配置 MPLS TE 的所有高级特性和应用的基础。

配置动态 MPLS TE 隧道需要在骨干网各节点上进行以下配置，其中配置链路的带宽和配置 MPLS TE 隧道的约束条件为可选步骤。

（1）使能 MPLS TE 和 RSVP-TE。

（2）配置 MPLS TE 隧道接口。

（3）（可选）配置链路的带宽。

本项配置任务与静态 MPLS TE 隧道中 5.6.3 节介绍的链路带宽的配置方法完全一样，参见即可。

（4）配置 TE 信息发布。

（5）（可选）配置 MPLS TE 隧道的约束条件。

（6）配置 CSPF 路径计算。

在配置动态 MPLS TE 隧道之前，需完成以下任务。

- 配置 IGP 路由协议，使各节点间的 IP 路由可达。
- 配置各 LSR 节点的 LSR-ID。
- 配置各 LSR 节点的全局和公网接口 MPLS 能力。

5.7.1 使能 MPLS TE 和 RSVP-TE

动态 MPLS TE 隧道中的这项配置任务与静态 MPLS TE 隧道配置任务中的 5.6.1 节介绍的使能MPLS TE的配置非常相似，只是多了一个在全局和公网接口上使能RSVP-TE 功能，具体配置步骤见表 5-16，需在 MPLS TE 隧道的各节点上进行配置。

表 5-16　使能 **MPLS TE** 和 **RSVP-TE** 的配置步骤

步骤	命令	说明
1	**system-view** 例如：<Huawei> **system-view**	进入系统视图
2	**mpls** 例如：[Huawei] **mpls**	使能全局 MPLS 功能，进入 MPLS 视图
3	**mpls te** 例如：[Huawei-mpls] **mpls te**	全局使能本节点的全局 MPLS TE 功能
4	**mpls rsvp-te** 例如：[Huawei-mpls] **mpls rsvp-te**	使能本节点 RSVP-TE。缺省情况下，全局 RSVP-TE 功能处于未使能状态，可用 **undo mpls rsvp-te** 命令去使能
5	**quit** 例如：[Huawei-mpls] **quit**	返回系统视图
6	**interface** *interface-type interface-number* 例如：[Huawei] **interface** gigabitethernet 1/0/0	进入 MPLS/IP 公网接口的接口视图
7	**mpls** 例如：[Huawei-GigabitEthernet1/0/0] **mpls**	在以上公网接口上使能 MPLS
8	**mpls te** 例如：[Huawei-GigabitEthernet1/0/0] **mpls te**	在以上公网接口上使能 MPLS TE
9	**mpls rsvp-te**	在以上公网接口上使能接口的 RSVP-TE。缺省情况下，接口下的 RSVP-TE 功能处于未使能状态，可用 **undo mpls rsvp-te** 命令去使能

5.7.2 配置 MPLS TE 隧道接口

动态 MPLS TE 隧道的这项配置任务也与静态 MPLS TE 隧道中的 5.6.2 节介绍的 MPLS TE 隧道接口的配置方法非常类似，只是这里要在隧道接口使能 RSVP-TE 信令协议，具体配置步骤见表 5-17。**仅需在 MPLS TE 隧道 Ingress 节点进行配置**。

表 5-17　**MPLS TE** 隧道接口的配置步骤

步骤	命令	说明
1	**system-view** 例如：<Huawei> **system-view**	进入系统视图
2	**interface tunnel** *interface-number* 例如：[Huawei] **interface tunnel** 0/0/1	创建 Tunnel 接口并进入 Tunnel 接口视图。其他说明参见 5.6.2 节表 5-14 中的第 2 步
3	**ip address** *ip-address* { *mask* | *mask-length* } [**sub**] 例如：[Huawei-Tunnel0/0/1] **ip address** 10.1.1.1 24	（二选一）配置 Tunnel 接口的 IP 地址
	ip address unnumbered interface *interface-type interface-number* 例如：[Huawei-Tunnel0/0/1] **ip address unnumbered interface** loopback 0	（二选一）配置隧道接口借用其他接口的 IP 地址。参数 *interface-type interface-number* 指定被借用的接口。其他说明参见 5.6.2 节表 5-14 的第 3 步
4	**tunnel-protocol mpls te** 例如：[Huawei-Tunnel0/0/1] **tunnel-protocol mpls te**	配置隧道协议为 MPLS TE。其他说明参见 5.6.2 节表 5-14 的第 4 步

（续表）

步骤	命令	说明
5	**destination** *dest-ip-address* 例如：[Huawei-Tunnel0/0/1] **destination** 4.4.4.9	配置隧道的目的 IP 地址。其他说明参见 5.6.2 节表 5-14 的第 5 步
6	**mpls te tunnel-id** *tunnel-id* 例如：[Huawei-Tunnel0/0/1] **mpls te tunnel-id** 100	配置隧道 ID。其他说明参见 5.6.2 节表 5-14 的第 6 步
7	**mpls te signal-protocol** rsvp-te 例如：[Huawei-Tunnel0/0/1] **mpls te signal-protocol rsvp-te**	配置隧道使用 RSVP-TE 作为信令协议。 缺省情况下，MPLS TE 建立隧道使用 RSVP-TE 信令协议，可用 **undo mpls te signal-protocol cr-static** 命令恢复缺省配置
8	**mpls te signalled tunnel-name** *tunnel-name* 例如：[Huawei-Tunnel0/0/1] **mpls te signalled tunnel-name** LSRAtoLSRC	（可选）配置 TE 隧道的名称。其他说明参见 5.6.2 节表 5-14 的第 6 步
9	**mpls te cspf disable** 例如：[Huawei-Tunnel0/0/1] **mpls te cspf disable**	（可选）使能在建立 MPLS TE 隧道时屏蔽 CSPF 计算 当全局使能了 CSPF 后，TE 隧道建立 LSP 时都会触发 CSPF 计算。但在 OptionC 方式跨域 VPN 场景下（参见《华为 MPLS VPN 学习指南》一书），由于两个域之间不会配置 IGP 协议，故不能生成 TEDB（流量工程数据库），导致 CSPF 算路不成功，无法建立跨域的 TE 隧道。此时，可以在配置的 TE 隧道接口下，**执行本命令，使当前 TE 隧道屏蔽 CSPF 算路功能，依靠直连路由或者静态路由建立跨域的 TE 隧道**。 【注意】执行本命令后，跳数限制、CSPF 的仲裁、SRLG 等 CR-LSP 的路径选择功能都将失效。 缺省情况下，未使能在 TE 隧道下建立 LSP 时屏蔽 CSPF 计算功能，可用 **undo mpls te cspf disable** 命令去使能在 TE 隧道下建立 LSP 时屏蔽 CSPF 计算功能
10	**mpls te commit** 例如：[Huawei-Tunnel0/0/1] **mpls te commit**	提交隧道当前配置。凡是发生了 MPLS 隧道配置更改都要执行本命令，使配置更改生效。 其他说明参见 5.6.2 节表 5-14 的第 9 步

5.7.3 配置 TE 信息发布

MPLS TE 隧道的路径计算是由 CSPF 完成的，计算时会考虑链路的带宽、颜色等 TE 链路属性。这些 TE 链路属性需要在 MPLS 区域中的各 LSR 间扩散并同步，最终形成一致的 TEDB，供给 CSPF 计算。目前设备支持两种 TE 信息发布来形成 TEDB。

（1）OSPF TE

OSPF TE 在 OSPF 原有协议基础上扩展使用 Opaque Type 10 LSA（包括 Type 1 和 Type 2 这两种类型 TLV 的 TE LSA）携带链路的 TE 属性信息，能在 MPLS 区域中的各 LSR 间扩散 TE 信息，形成 TEDB 提供给 CSPF 计算。有关 OSPF TE 详细内容参见本章 5.2.2 节。

使用 OSPF TE 配置 TE 信息发布的方法见表 5-18，**需在 MPLS TE 隧道各节点上进行配置**。缺省情况下，OSPF 区域不支持 TE。因此，必须使能 OSPF 的 Opaque，并且只有当至少有一个邻居处于 FULL 状态时，才会产生 Opaque Type 10 LSA。

表 5-18　　使用 **OSPF TE** 发布 **TE** 信息发布的配置步骤

步骤	命令	说明
1	**system-view** 例如：<Huawei> **system-view**	进入系统视图
2	**ospf** [*process-id*] 例如：[Huawei] **ospf**	进入 OSPF 视图
3	**opaque-capability enable** 例如：[Huawei-ospf-1] **opaque-capability enable**	使能 OSPF 的 Opaque 能力，从而 OSPF 进程可以生成 Opaque LSA，并能从邻居设备接收 Opaque LSA。 缺省情况下，禁止 opaque-lsa 能力，可用 **undo opaque-capability** 命令禁止对 Opaque LSA 进行操作
4	**area** *area-id* 例如：[Huawei-ospf-1] **area** 0	进入 OSPF 的区域视图
5	**mpls-te enable** [**standard-complying**] 例如：[Huawei-ospf-1] **mpls-te enable**	在当前 OSPF 区域使能 TE。可选项 **standard-complying** 用来指定只接收标准格式的 LSA，即若 TE LSA 中有超过一个的 Top level TLV，则认为该 LSA 错误。 缺省情况下，OSPF 区域不支持 TE，可用 **undo mpls-te** 命令取消当前 OSPF 区域的 MPLS TE 特性

（2）IS-IS TE

IS-IS TE 是 IS-IS 为了支持 MPLS TE 而做的扩展，它通过在 IS-IS LSP 报文中定义新的 TLV（包括 Type 22 和 Type 135 这两种 TLV）的方式，携带该设备 MPLS TE 的配置信息，通过 LSP 的泛洪同步，实现各 LSR 间 MPLS TE 信息的泛洪和同步。

IS-IS TE 把所有 LSP 中携带的 TE 信息提取出来，传递给 MPLS 的 CSPF 模块，用来计算隧道路径。有关 IS-IS TE 详细内容参见 5.2.3 节。使用 IS-IS TE 配置 TE 信息发布的方法见表 5-19，**需在 MPLS TE 隧道各节点上进行配置**。

表 5-19　　使用 **IS-IS TE** 发布 **TE** 信息发布的配置步骤

步骤	命令	说明
1	**system-view** 例如：<Huawei> **system-view**	进入系统视图
2	**isis** [*process-id*] 例如：[Huawei] **isis**	进入 IS-IS 协议视图
3	**cost-style** { **compatible** [**relax-spf-limit**] | **wide** | **wide-compatible** } 例如：[Huawei-isis-1] **cost-style narrow-compatible**	配置 IS-IS 的 Wide Metric 属性。命令中的选项说明如下。 • **compatible**：多选一选项，定 IS-IS 设备可以接收和发送开销类型为 narrow 和 wide 的路由。 • **relax-spf-limit**：可选项，指定 IS-IS 设备可以接收开销值大于 1023 的路由。如果不设置本选项。 　• 当路由开销值小于等于 1023，且该路由经过的所有接口的链路开销值都小于等于 63 时，则这条路由的开销值按照实际值接收，即路由的开销值为该路由所经过的所有接口的链路开销总和

（续表）

步骤	命令	说明
3	**cost-style** { **compatible** [**relax-spf-limit**] | **wide** | **wide-compatible** } 例如：[Huawei-isis-1] **cost-style narrow-compatible**	• 当路由开销值小于等于 1023，但该路由经过的所有接口中有的接口链路开销值大于 63 时，则设备只能学到该接口所在设备的其他接口的直连路由和该接口所引入的路由，路由的开销值按照实际值接收，路由此后要经过的接口将丢弃该路由。 • 当路由开销值大于 1023，设备可以接收链路开销值小于 1023 的接口所在网段的所有路由时，如果路由开销值大于 1023 按照 1023 接收。 • **wide**：多选一选项，指定 IS-IS 设备只能接收和发送开销类型为 wide 的路由。wide 模式下路由的开销值取值范围是 1～16777215。 • **wide-compatible**：多选一选项，指定 IS-IS 设备可以接收开销类型为 narrow 和 wide 的路由，但却只发送开销类型为 wide 的路由。 【说明】IS-IS TE 扩展使用 IS 可达性 TLV（22）的子 TLV 携带 TE 属性信息。因此，必须使能 IS-IS 的 WideMetric 特性，可以设置为 wide、compatible 或 wide-compatible 属性 缺省情况下，IS-IS 只收发 narrow 方式表示路由权值的报文，可用 **undo cost-style** 命令恢复 IS-IS 设备接收和发送路由开销类型为缺省类型
4	**traffic-eng** [**level-1** | **level-2** | **level-1-2**] 例如：[Huawei-isis-1] **traffic-eng level-1**	使能 IS-IS 进程不同层次的 TE 特性，具体要根据设备所处的 IS-IS 层次来选择。命令中的选项说明如下。 • **level-1**：多选一选项，设置 Level-1 的 IS-IS TE。 • **level-2**：多选一选项，设置 Level-2 的 IS-IS TE。 • **level-1-2**：多选一选项，设置 Level-1-2 的 IS-IS TE。 如果在使能 IS-IS TE 时不指定 Level，则同时对 Level-1 和 Level-2 生效。 缺省情况下，IS-IS 进程不支持 TE，可用 **undo traffic-eng** [**level-1** | **level-2** | **level-1-2**] 命令去使能指定层次的 TE 特性
5	**te-set-subtlv** { **bw-constraint** *bw-constraint-value* | **lo-multiplier** *lo-multiplier-value* | **unreserve-bw-sub-pool** *unreserve-bw-sub-pool-value* } * 例如：[Huawei-isis-1] **te-set-subtlv bw-constraint** 200 **lo-multiplier** 201 **unreserve-bw-sub-pool** 202	（可选）设置携带 DS-TE 参数的子 TLV 的类型。命令中的参数说明如下。 • **bw-constraint** *bw-constraint-value*：可多选参数，指定带宽约束（BW-Constraint）的子 TLV 值，整数形式，取值范围是 19～254。 • **lo-multiplier** *lo-multiplier-value*：可多选参数，指定本地过预订倍数 LOM（Local Overbooking Multipliers）的子 TLV 值，整数形式，取值范围是 19～254。 • **unreserve-bw-sub-pool** *unreserve-bw-sub-pool-value*：可多选参数，指定子池未预订带宽（Unreserve-BW-Sub-Pool）的子 TLV 值，整数形式，取值范围是 19～254

（续表）

步骤	命令	说明
5	**te-set-subtlv** { **bw-constraint** *bw-constraint-value* | **lo-multiplier** *lo-multiplier-value* | **unreserve-bw-sub-pool** *unreserve-bw-sub-pool-value* } * 例如：[Huawei-isis-1] **te-set-subtlv bw-constraint** 200 **lo-multiplier** 201 **unreserve-bw-sub-pool** 202	【说明】由于 DS-TE 参数的各子 TLV 的类型尚未形成标准，当使用不同厂商提供的设备互联时，用户需要手工配置这些子 TLV 值，并使他们的值保持一致。配置后，TEDB 将重新生成，进而导致 TE 隧道重建。如果全是华为设备，则可直接采用缺省取值。 缺省情况下，带宽约束的子 TLV 为 252；本地过预订倍数 LOM 的子 TLV 为 253；子池未预订带宽的子 TLV 为 251，可用 **undo te-set-subtlv** { **bw-constraint** [*bw-constraint-value*] | **lo-multiplier** [*lo-multiplier-value*] | **unreserve-bw-sub-pool** [*unreserve-bw-sub-pool-value*] } * 命令恢复缺省设置

5.7.4 配置 MPLS TE 隧道的约束条件

在入节点上配置隧道的显式路径等约束条件，可以精确、灵活地控制 RSVP-TE 隧道的建立。本项配置任务包括两个方面。

（1）配置 MPLS TE 显式路径

如果需要配置隧道的显式路径约束条件时，需要先创建显式路径。显式路径由一系列节点构成，按配置的先后顺序组成一条向量路径。显式路径中的 IP 地址是指节点上接口的 IP 地址，通常采用出节点上 Loopback 接口的地址作为显式路径的目的地址。通过配置显式路径，可以指定 CR-LSP 必须经过某些路径或节点，更好地进行资源的合理分配，增加隧道路径的可控性。

显式路径上的两个相邻节点之间存在两种关系。

- 严格下一跳（strict）：两个节点必须直接相连，用于精确控制 LSP 所经过的路径。
- 松散下一跳（loose）：两个节点之间可以存在其他节点。

严格方式与松散方式可以单独使用，也可以混合使用。

（2）配置 MPLS TE 隧道的约束条件

指定隧道的约束条件，CSPF 会根据隧道上配置的约束条件进行路径计算，保证 CR-LSP 的正确建立。

以上两方面的具体配置步骤见表 5-20。

表 5-20　　MPLS TE 隧道约束条件的配置步骤

步骤	命令	说明
1	**system-view** 例如：<Huawei> **system-view**	进入系统视图
2	**explicit-path** *path-name* 例如：[Huawei] **explicit-path** p1	创建显式路径，进入显式路径视图。参数 *path-name* 用来指定隧道的显式路径名称，字符串形式，不区分大小写，不支持空格，长度范围为大于等于 1。 【注意】必须启动 MPLS TE 功能后才能配置隧道的显式路径。且显式路径上的节点地址不能重复，也不能形成环路。如果有环路，CSPF 将检测出环路，无法成功计算出路径

（续表）

<table>
<tr><th>步骤</th><th>命令</th><th>说明</th></tr>
<tr><td>2</td><td>explicit-path path-name
例如：[Huawei] explicit-path p1</td><td>缺省情况下，没有配置隧道的显式路径，可用 undo explicit-path path-name 命令删除配置的指定显式路径</td></tr>
<tr><td>3</td><td>next hop ip-address [include [[loose | strict] | [incoming | outgoing]] * | exclude]
例如：[Huawei-explicit-path-p1] next hop 10.0.0.125 exclude</td><td>指定显式路径的下一个节点。命令中的参数和选项说明如下。
• ip-address：指定显式路径中的下一个节点 IP 地址。
• include [[loose | strict] | [incoming | outgoing]] *：二选一可选项，指定在显式路径中包含此节点。其中的选项如下。
　• strict：表示严格显式路径，参数 ip-address 指定的节点与本节点必须直连。缺省情况下，采用 strict 模式，即加入的下一跳与上一节点必须是直连的。作为一种约束条件，显式路径可以指定经过或者不经过某些节点。
　• loose：表示松散显式路径，参数 ip-address 指定的节点与本节点可以不是直连的。
　• incoming：指定参数 ip-address 为当前配置的下一个节点的入接口地址。
　• outgoing：指定参数 ip-address 为当前配置的下一个节点的出接口地址。
　• exclude：二选一可选项，指定显式路径不能经过参数 ip-address 指定的节点。
【说明】需通过本命令依次把路径中的每个下一跳列出来，构建完整的显式路径。如果指定的 ip-address 是当前配置下一个节点的入接口地址，建议配置 incoming 参数；如果指定的 ip-address 是当前配置下一个节点的出接口地址，建议配置 outgoing 参数。
缺省情况下，没有在显式路径中指定下一个节点，可用 undo next hop ip-address 命令删除指定的下一跳</td></tr>
<tr><td colspan="3">执行以下命令增加、修改或删除显式路径中的节点</td></tr>
<tr><td rowspan="2">4</td><td>list hop [ip-address]
例如：[Huawei-explicit-path-path1] list hop</td><td>查看显式路径节点信息。可选参数 ip-address 用来指定要查看当前显式路径配置的节点的 IP 地址。如果不指定本参数，则查看当前显式路径下的所有节点</td></tr>
<tr><td>add hop ip-address1 [include [[loose | strict] | [incoming | outgoing]] * | exclude] { after | before } ip-address2
例如: [Huawei-explicit-path-p1] add hop 10.2.2.2 exclude after 10.1.1.1</td><td>（可选）向显式路径中插入一个节点。本命令中的大多数参数和选项与第 3 步中的命令中的参数和选项一样，只不过这里是插入节点的操作，下面仅介绍不同的参数和选项。
• after：二选一选项，表示在参数 ip-address2 后插入参数 ip-address1 指定的节点。
• before：二选一选项，表示在参数 ip-address2 前插入参数 ip-address1 指定的节点。
• ip-address2：指定已经在显式路径中的节点接口 IP 地址或节点 Router ID。
如果指定的 ip-address1 是新增节点的入接口地址，建议配置 incoming 参数；如果指定的 ip-address1 是新增节点的出接口地址，建议配置 outgoing 参数</td></tr>
</table>

（续表）

步骤	命令	说明
4	**modify hop** *ip-address1 ip-address2* [**include** [[**loose** | **strict**] | [**incoming** | **outgoing**]] * | **exclude**] 例如：[Huawei-explicit-path-p1] **modify hop** 1.1.1.9 2.2.2.9	（可选）修改显式路径中的节点地址。参数 *ip-address1 ip-address2* 指定将显式路径中的IP地址 *ip-address1* 修改为 *ip-address2*。其他选项与本表第3步中的对应选项作用一样。 【说明】如果指定的 *ip-address2* 是修改后节点的入接口地址，建议配置 **incoming** 参数；如果指定的 *ip-address2* 是修改后节点的出接口地址，建议配置 **outgoing** 参数
	delete hop *ip-address* 例如：[Huawei-explicit-path-p1] **delete hop** 10.10.10.10	（可选）从显式路径中删除一个节点。参数 *ip-address* 用来指定要删除节点的IP地址。此节点必须是显式路径中存在的节点
5	**quit** 例如：[Huawei-explicit-path-p1] **quit**	返回系统视图
6	**interface tunnel** *tunnel-number* 例如：[Huawei] **interface tunnel** 0/0/1	进入MPLS TE隧道的Tunnel接口视图
7	**mpls te path explicit-path** *path-name* 例如：[Huawei-Tunnel0/0/1] **mpls te path explicit-path** p1	配置隧道应用的显式路径，该路径是在本表第2步创建的显式路径。缺省情况下，没有为当前隧道配置显式路径，**undo mpls te path explicit-path** *path-name* 命令用来删除显式路径
8	**mpls te commit** 例如：[Huawei-Tunnel0/0/1] **mpls te commit**	提交隧道当前配置，使以上配置生效

5.7.5 配置MPLS TE路径计算

为了计算出满足指定约束条件的隧道路径，**需要在隧道的入节点上配置 CSPF**。他在计算路径时将考虑以下条件。

- IGP-TE（如OSPF TE或IS-IS TE）TEDB中维护的链路状态信息。
- IGP-TE TEDB中与网络资源状态相关的属性（链路最大带宽、最大可预留带宽、亲和属性等）。
- 由用户指定的路径约束条件（显式路径）。

路径计算的配置步骤见表5-21，需在入节点上配置。

说明　隧道入节点不使能CSPF时，RSVP-TE隧道也可以建立成功。但是为了使隧道路径能够满足预设的约束条件，建议使能CSPF。推荐在所有的Transit节点也使能CSPF。

表5-21　MPLS TE路径计算的配置步骤

步骤	命令	说明
1	**system-view** 例如：<Huawei> **system-view**	进入系统视图
2	**mpls** 例如：[Huawei] **mpls**	进入MPLS视图

（续表）

步骤	命令	说明
3	**mpls te cspf** 例如：[Huawei-mpls] **mpls te cspf**	使能本节点的 CSPF 功能。在使能 CSPF 前，应先在 MPLS 视图下使能 MPLS TE。缺省情况下，未使能 CSPF，可用 **undo mpls te cspf** 命令去使能 CSPF
4	**mpls te cspf preferred-igp** { **isis** [*isis-process-id* [**level-1** | **level-2**]] | **ospf** [*ospf-process-id* [**area** { *area-id-1* | *area-id-2* }]] } 例如：[Huawei-mpls] **mpls te cspf preferred-igp ospf**	（可选）配置 CSPF 选路时的首选 IGP 协议，以允许用户设置 CSPF 优先采用 IS-IS 协议生成的 TEDB 来计算 CR-LSP 路径。如果根据 IS-IS 的 TEDB 数据计算失败，才会进一步采用 OSPF 的 TEDB 数据再次计算。**如果骨干网只配置了单一的 IGP 协议来发布 TE 信息（OSPF TE 或者 IS-IS TE），则不需要执行该步骤**。命令中的参数和选项说明如下。 • **isis**：二选一选项，指定优先选择 IS-IS 的 TEDB 进行选路。 • *isis-process-id*：可选参数，指定 IS-IS 进程号，整数形式，取值范围是 1～65535，缺省为 1。 • **level-1** | **level-2**：可选项，指定 CSPF 在计算时优先选择 level-1 或 level-2 的数据。 • **ospf**：二选一选项，指定优先选择 OSPF 的 TEDB 选路。 • *ospf-process-id*：可选参数，指定 OSPF 进程号，整数形式，取值范围是 1～65535，缺省为 1。 • *area-id-1* | *area-id-2*：可选参数，指定优选的 OSPF 区域（以编号形式或 IP 地址形式）。 【说明】CSPF 在进行选路计算时，缺省优先采用 OSPF 协议生成的 TEDB 来计算 CR-LSP 路径。如果根据 OSPF 的 TEDB 能够计算出路径，则不会再根据 IS-IS 协议的 TEDB 数据计算；如果根据 OSPF 的 TEDB 数据计算失败，则会根据 IS-IS 协议的 TEDB 数据再次计算，可用 **undo mpls te cspf preferred-igp** 命令恢复缺省设置

5.7.6　动态 MPLS TE 隧道配置管理

当完成动态 MPLS TE 隧道的所有配置后，可执行以下 display 命令查看相关配置，验证配置结果。

■ **display mpls te link-administration bandwidth-allocation** [**interface** *interface-type interface-number*]：查看指定或所有链路带宽分配信息。

■ **display ospf** [*process-id*] **mpls-te** [**area** *area-id*] [**self-originated**]：查看指定的 OSPF TE 信息。

■ 执行以下命令查看 IS-IS TE 状态：

- **display isis traffic-eng advertisements.**
- **display isis traffic-eng link.**
- **display isis traffic-eng network.**
- **display isis traffic-eng statistics.**
- **display isis traffic-eng sub-tlvs.**

■ **display explicit-path** [[**name**] *path-name*] [**tunnel-interface** | **verbose**]：查看指

定或所有已经配置的显式路径。

■ **display mpls te cspf destination** *ip-address* [**affinity** *properties* [**mask** *mask-value*] | **bandwidth** { **ct0** *ct0-bandwidth* | **ct1** *ct1-bandwidth* | **ct2** *ct2-bandwidth* | **ct3** *ct3-bandwidth* | **ct4***ct4-bandwidth* | **ct5** *ct5-bandwidth* | **ct6** *ct6-bandwidth* | **ct7** *ct7-bandwidth* } * | **explicit-path** *path-name* | **hop-limit** *hop-limit-number* | **metric-type** { **igp** | **te** } | **priority** *setup-priority* | **srlg-strict***exclude-path-name* | **tie-breaking** { **random** | **most-fill** | **least-fill** }] * [**hot-standby** [**explicit-path** *path-name* | **overlap-path** | **affinity** *properties* [**mask** *mask-value*] | **hop-limit** *hop-limit-number* | **srlg** { **preferred** | **strict** }] *]：查看满足指定条件的 CSPF 计算的路径。

■ **display mpls te cspf tedb** { **all** | **area** { *area-id* | *area-id-ip* } | **interface** *ip-address* | **network-lsa** | **node** [*router-id*] | **srlg** *srlg-number* | **overload-node** }：查看满足指定条件的用于 CSPF 计算的 TEDB 信息。

■ **display mpls rsvp-te**：查看 RSVP 的相关信息。

■ **display mpls rsvp-te established** [**interface** *interface-type interface-number peer-ip-address*]：查看已建立的 RSVP LSP 信息。

■ **display mpls rsvp-te peer** [**interface** *interface-type interface-number*]：查看 RSVP 邻居参数。

■ **display mpls rsvp-te reservation** [**interface** *interface-type interface-number peer-ip-address*]：查看指定或所有接口上的 RSVP 资源预留信息。

■ **display mpls rsvp-te request** [**interface** *interface-type interface-number peer-ip-address*]：查看指定或接口上的 RSVP-TE 请求消息信息。

■ **display mpls rsvp-te sender** [**interface** *interface-type interface-number peer-ip-address*]：查看指定或所有接口上的 RSVP 发送方信息。

■ **display mpls rsvp-te statistics** { **global** | **interface** [*interface-type interface-number*] }：查看指定或所有接口上的 RSVP-TE 运行统计信息。

■ **display mpls te link-administration admission-control** [**interface** *interface-type interface-number* | **stale-interface** *interface-index*]：查看本地接纳的隧道。

■ **display mpls te tunnel** [**destination** *ip-address*] [**lsp-id** *ingress-lsr-id session-id local-lsp-id*] [**lsr-role** { **all** | **egress** | **ingress** | **remote** | **transit** }] [**name** *tunnel-name*] [{ **incoming-interface** | **interface** | **outgoing-interface** } *interface-type interface-number*] [**te-class0** | **te-class1** | **te-class2** | **te-class3** | **te-class4** | **te-class5** | **te-class6** | **te-class7**] [**verbose**]：查看隧道信息。

■ **display mpls te tunnel statistics** 或者 **display mpls lsp statistics**：查看隧道统计信息。

■ **display mpls te tunnel-interface** [**tunnel** *interface-number* | **auto-bypass-tunnel** [*tunnel-name*]]：查看指定或所有 MPLS TE 隧道的接口信息。

■ **display mpls te tunnel c-hop** [*tunnel-name*] [**lsp-id** *ingress-lsr-id session-id lsp-id*]：查看指定或所有隧道选路结果。

■ **display mpls te session-entry** [*ingress-lsr-id tunnel-id egress-lsr-id*]：查看指定或所有隧道的 LSP 会话详细信息。

5.7.7 动态 MPLS TE 隧道配置示例

如图 5-21 所示，某企业自建 MPLS 骨干网，LSRA、LSRB、LSRC 均属于 MPLS 骨干网设备。MPLS 骨干网的路由协议使用 IS-IS，都属于 Level-2。现要求在 MPLS 骨干网创建公网隧道承载 L2VPN 或 L3VPN 业务，同时要求该隧道适应网络拓扑变化保证数据传输的稳定性。

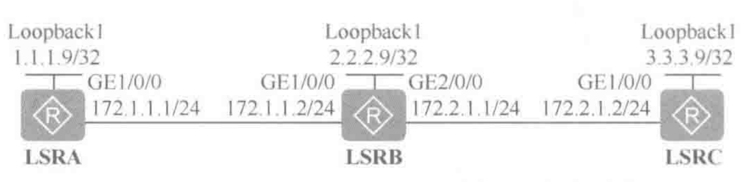


图 5-21　动态 MPLS TE 隧道配置示例拓扑结构

为了实现该需求，需要使用 RSVP-TE 信令建立一条动态 MPLS TE 隧道。

1. 基本配置思路分析

本示例采用基于 RSVP 信令协议的动态 MPLS TE 隧道来实现。根据前面介绍的配置任务，再结合本示例的具体要求可以得出以下基本配置思路。

（1）配置各设备各接口 IP 地址及 IS-IS 路由，实现 MPLS 骨干网设备路由可达。

（2）在各节点全局和公网接口上使能 MPLS TE 和 RSVP-TE 功能，以便建立 MPLS TE 隧道。

（3）使能 IS-IS TE，配置 IS-IS 的 Wide Metric 属性，使得 TE 信息可以通过 IS-IS 发布到其他节点。

（4）在入节点创建隧道接口，配置隧道属性，使能 MPLS TE CSPF，创建动态 MPLS TE 隧道。如果要双向通信的话，则需要创建两条相反方向的 MPLS TE 隧道。

2. 具体配置步骤

（1）配置各接口（包括 Loopback 接口）的 IP 地址和 IS-IS 路由。假设区域 ID 为 5，系统 ID 从 1 开始依次分配，都位于 Level-2 层次。

\# LSRA 上的配置。

```
<Huawei> system-view
[Huawei] sysname LSRA
[LSRA] interface gigabitethernet 1/0/0
[LSRA-GigabitEthernet1/0/0] ip address 172.1.1.1 255.255.255.0
[LSRA-GigabitEthernet1/0/0] quit
[LSRA] interface loopback 1
[LSRA-LoopBack1] ip address 1.1.1.9 255.255.255.255
[LSRA-LoopBack1] quit
[LSRA] isis 1
[LSRA-isis-1] network-entity 00.0005.0000.0000.0001.00  #---配置 NET 实体名称，区域 ID 为 5，系统 ID 为 1
[LSRA-isis-1] is-level level-2  #---指定当前路由器级别为 Level-2
[LSRA-isis-1] quit
[LSRA] interface gigabitethernet 1/0/0
[LSRA-GigabitEthernet1/0/0] isis enable 1  #---在以上接口上启动 IS-IS 路由进程 1
[LSRA-GigabitEthernet1/0/0] quit
[LSRA] interface loopback 1
[LSRA-LoopBack1] isis enable 1
[LSRA-LoopBack1] quit
```

LSRB 上的配置。

```
<Huawei> system-view
[Huawei] sysname LSRB
[LSRB] interface gigabitethernet 1/0/0
[LSRB-GigabitEthernet1/0/0] ip address 172.1.1.2 255.255.255.0
[LSRB-GigabitEthernet1/0/0] quit
[LSRB] interface gigabitethernet 2/0/0
[LSRB-GigabitEthernet2/0/0] ip address 172.2.1.1 255.255.255.0
[LSRB-GigabitEthernet2/0/0] quit
[LSRB] interface loopback 1
[LSRB-LoopBack1] ip address 2.2.2.9 255.255.255.255
[LSRB-LoopBack1] quit
[LSRB] isis 1
[LSRB-isis-1] network-entity 00.0005.0000.0000.0002.00
[LSRB-isis-1] is-level level-2
[LSRB-isis-1] quit
[LSRB] interface gigabitethernet 1/0/0
[LSRB-GigabitEthernet1/0/0] isis enable 1
[LSRB-GigabitEthernet1/0/0] quit
[LSRB] interface gigabitethernet 2/0/0
[LSRB-GigabitEthernet2/0/0] isis enable 1
[LSRB-GigabitEthernet2/0/0] quit
[LSRB] interface loopback 1
[LSRB-LoopBack1] isis enable 1
[LSRB-LoopBack1] quit
```

LSRC 上的配置。

```
<Huawei> system-view
[Huawei] sysname LSRC
[LSRC] interface gigabitethernet 1/0/0
[LSRC-GigabitEthernet1/0/0] ip address 172.2.1.2 255.255.255.0
[LSRC-GigabitEthernet1/0/0] quit
[LSRC] interface loopback 1
[LSRC-LoopBack1] ip address 3.3.3.9 255.255.255.255
[LSRC-LoopBack1] quit
[LSRC] isis 1
[LSRC-isis-1] network-entity 00.0005.0000.0000.0003.00
[LSRC-isis-1] is-level level-2
[LSRC-isis-1] quit
[LSRC] interface gigabitethernet 1/0/0
[LSRC-GigabitEthernet1/0/0] isis enable 1
[LSRC-GigabitEthernet1/0/0] quit
[LSRC] interface loopback 1
[LSRC-LoopBack1] isis enable 1
[LSRC-LoopBack1] quit
```

以上配置完成后，在各节点上执行 **display ip routing-table** 命令，可以看到各节点间已相互都学到了到对方的路由。以下是在 LSRA 上执行该命令的输出示例，参见输出信息中的粗体字部分。

```
[LSRA] display ip routing-table
Route Flags: R - relay, D - download to fib
------------------------------------------------------------------------------
Routing Tables: Public
         Destinations : 11          Routes : 11
```

```
Destination/Mask      Proto    Pre  Cost       Flags NextHop         Interface

        1.1.1.9/32    Direct   0    0            D   127.0.0.1       LoopBack1
        2.2.2.9/32    ISIS-L2  15   10           D   172.1.1.2       GigabitEthernet1/0/0
        3.3.3.9/32    ISIS-L2  15   20           D   172.1.1.2       GigabitEthernet1/0/0
       127.0.0.0/8    Direct   0    0            D   127.0.0.1       InLoopBack0
      127.0.0.1/32    Direct   0    0            D   127.0.0.1       InLoopBack0
127.255.255.255/32    Direct   0    0            D   127.0.0.1       InLoopBack0
      172.1.1.0/24    Direct   0    0            D   172.1.1.1       GigabitEthernet1/0/0
      172.1.1.1/32    Direct   0    0            D   127.0.0.1       GigabitEthernet1/0/0
    172.1.1.255/32    Direct   0    0            D   127.0.0.1       GigabitEthernet1/0/0
      172.2.1.0/24    ISIS-L2  15   20           D   172.1.1.2       GigabitEthernet1/0/0
255.255.255.255/32    Direct   0    0            D   127.0.0.1       InLoopBack0
```

（2）配置 MPLS 基本能力，使能 MPLS TE、RSVP-TE。

在各节点全局使能 MPLS、MPLS TE 和 RSVP-TE，在隧道沿途的公网接口上使能 MPLS、MPLS TE 和 RSVP-TE。

LSRA 上的配置。

```
[LSRA] mpls lsr-id 1.1.1.9
[LSRA] mpls
[LSRA-mpls] mpls te
[LSRA-mpls] mpls rsvp-te
[LSRA-mpls] quit
[LSRA] interface gigabitethernet 1/0/0
[LSRA-GigabitEthernet1/0/0] mpls
[LSRA-GigabitEthernet1/0/0] mpls te
[LSRA-GigabitEthernet1/0/0] mpls rsvp-te
[LSRA-GigabitEthernet1/0/0] quit
```

LSRB 上的配置。

```
[LSRB] mpls lsr-id 2.2.2.9
[LSRB] mpls
[LSRB-mpls] mpls te
[LSRB-mpls] mpls rsvp-te
[LSRB-mpls] quit
[LSRB] interface gigabitethernet 1/0/0
[LSRB-GigabitEthernet1/0/0] mpls
[LSRB-GigabitEthernet1/0/0] mpls te
[LSRB-GigabitEthernet1/0/0] mpls rsvp-te
[LSRB-GigabitEthernet1/0/0] quit
[LSRB] interface gigabitethernet 2/0/0
[LSRB-GigabitEthernet2/0/0] mpls
[LSRB-GigabitEthernet2/0/0] mpls te
[LSRB-GigabitEthernet2/0/0] mpls rsvp-te
[LSRB-GigabitEthernet2/0/0] quit
```

LSRC 上的配置。

```
[LSRC] mpls lsr-id 3.3.3.9
[LSRC] mpls
[LSRC-mpls] mpls te
[LSRC-mpls] mpls rsvp-te
[LSRC-mpls] quit
[LSRC] interface gigabitethernet 1/0/0
[LSRC-GigabitEthernet1/0/0] mpls
[LSRC-GigabitEthernet1/0/0] mpls te
```

```
[LSRC-GigabitEthernet1/0/0] mpls rsvp-te
[LSRC-GigabitEthernet1/0/0] quit
```

（3）配置 IS-IS TE，包括修改 IS-IS 的 Wide Metric 属性（可以是 compatible、wide 或 wide-compatible 类型），在设备所处的 Level-2 层次使能 IS-IS TE。

LSRA 上的配置。

```
[LSRA] isis 1
[LSRA-isis-1] cost-style wide   #---设置 ISOIS 开销为 wide 类型
[LSRA-isis-1] traffic-eng level-2   #---在 Level-2 级别使能 IS-IS TE，因为在前面已把各节点配置为 Level-2 级别
[LSRA-isis-1] quit
```

LSRB 上的配置。

```
[LSRB] isis 1
[LSRB-isis-1] cost-style wide
[LSRB-isis-1] traffic-eng level-2
[LSRB-isis-1] quit
```

LSRC 上的配置。

```
[LSRC] isis 1
[LSRC-isis-1] cost-style wide
[LSRC-isis-1] traffic-eng level-2
[LSRC-isis-1] quit
```

（4）配置 MPLS TE 隧道接口，使能 MPLS TE CSPF。

在隧道入节点上创建 Tunnel 接口，并配置 Tunnel 接口的 IP 地址、隧道协议、目的 IP 地址、Tunnel ID、动态信令协议，并执行 **mpls te commit** 命令使配置生效。

在 LSRA 上配置由 LSRA 到 LSRC 方向的 MPLS TE 隧道。

```
[LSRA] mpls
[LSRA-mpls] mpls te cspf   #---使能 CSPF 功能
[LSRA-mpls] quit
[LSRA] interface tunnel 0/0/1
[LSRA-Tunnel0/0/1] ip address unnumbered interface loopback 1
[LSRA-Tunnel0/0/1] tunnel-protocol mpls te   #---配置以上 Tunnel 接口封装 MPLS TE 协议
[LSRA-Tunnel0/0/1] destination 3.3.3.9   #---指定隧道目的端 IP 地址为 3.3.3.9，即 LSRC 的 Loopback1 接口 IP 地址
[LSRA-Tunnel0/0/1] mpls te tunnel-id 100   #---配置 MPLS TE 隧道 ID 为 100
[LSRA-Tunnel0/0/1] mpls te commit
[LSRA-Tunnel0/0/1] quit
```

在 LSRC 上配置由 LSRC 到 LSRA 方向的 MPLS TE 隧道。

```
[LSRC] mpls
[LSRC-mpls] mpls te cspf
[LSRC-mpls] quit
[LSRC] interface tunnel 0/0/1
[LSRC-Tunnel0/0/1] ip address unnumbered interface loopback 1
[LSRC-Tunnel0/0/1] tunnel-protocol mpls te
[LSRC-Tunnel0/0/1] destination 1.1.1.9
[LSRC-Tunnel0/0/1] mpls te tunnel-id 200
[LSRC-Tunnel0/0/1] mpls te commit
[LSRC-Tunnel0/0/1] quit
```

3. 配置结果验证

以上配置完成后，在 LSRA 或在 LSRC 上执行 **display interface tunnel** 命令可以看到在它们上面创建的隧道接口状态为 Up，表示 MPLS TE 隧道创建成功，前面的配置是正确的。以下是在 LSRA 上执行该命令的输出示例。

```
[LSRA] display interface tunnel
Tunnel0/0/1 current state : Up
```

```
Line protocol current state : Up
Last line protocol Up time : 2017-06-14 09:18:46
Description:
...
```

在 LSRA 或 LSRC 上执行 **display mpls te tunnel-interface** 命令可以看到在它们上面创建的隧道接口的基本信息，包括隧道 ID、入节点、出节点、状态等。以下是在 LSRA 上执行该命令的输出示例。

```
[LSRA] display mpls te tunnel-interface
    ----------------------------------------------------------------
                                   Tunnel0/0/1
    ----------------------------------------------------------------
    Tunnel State Desc      :  Up
    Active LSP              :  Primary LSP
    Session ID              :  100
    Ingress LSR ID          :  1.1.1.9          Egress LSR ID:   3.3.3.9
    Admin State             :  Up                 Oper State    :  Up
    Primary LSP State         : Up
      Main LSP State           : READY                    LSP ID   : 3
```

如果想要详细了解所创建的 MPLS TE 隧道信息，还可在 LSRA 或 LSRC 上执行 **display mpls te tunnel verbose** 命令查看。

在 LSRA 上执行 **display mpls te cspf tedb all** 命令查看 TEDB 中的链路信息。从中可以看出它包括了全部的三台设备，表示通过 CSPF 已在这三台设备上全部启用了。

```
[LSRA] display mpls te cspf tedb all
Maximum Nodes Supported: 1024     Current Total Node Number: 3
Maximum Links Supported: 2048     Current Total Link Number: 4
Maximum SRLGs supported: 5120     Current Total SRLG Number: 0
ID      Router-ID          IGP       Process-ID      Area             Link-Count
1       2.2.2.9            ISIS      1               Level-2          2
2       1.1.1.9            ISIS      1               Level-2          1
3       3.3.3.9            ISIS      1               Level-2          1
```

5.8 配置流量引入 MPLS TE 隧道

MPLS TE 隧道建立之后还不能自动引入流量，需要通过一定的方式将流量引入隧道，就像我们在《华为 VPN 学习指南》中介绍的 IPSec VPN、GRE VPN 等一样。

将用户流量引入到 MPLS TE 隧道的方式有四种：静态路由、策略路由、隧道策略和自动路由，参见 5.5 节。用户根据网络规划从中选择一种即可，推荐使用配置自动路由。在配置流量引入 MPLS TE 隧道之前，需完成静态/动态 MPLS TE 隧道，或静态/动态 DS-TE 隧道配置。有关静态/动态 DS-TE 隧道的配置方法参见本书第 9 章。

5.8.1 自动路由配置与管理

自动路由方式是指将 TE 隧道看作逻辑链路，与网络中的物理链路一起参与 IGP 路由计算，确定选择 TE 隧道作为某类流量的 IGP 路由路径时将使用对应的隧道接口作为路由出接口。根据某节点设备是否将 LSP 链路发布给邻居节点用于指导报文转发，自动路由又有两种配置方式。

■ 配置转发捷径：不将 TE 隧道发布给邻居节点，TE 隧道只能参与本地的路由计算，其他节点不能使用此隧道。

■ 配置转发邻接：将 TE 隧道发布给邻居节点，可使 TE 隧道参与全局的路由计算，其他节点也能使用此隧道。这是一种常用方式。

可根据实际情况选择其中一种方式，但仅需在隧道入节点上配置。在 MPLS TE 隧道**入节点**配置转发捷径的步骤见表 5-22，在 MPLS TE 隧道**入节点**配置转发邻接的步骤见表 5-23。

说明 转发捷径和转发邻接配置互斥，不能同时配置。配置转发邻接时，由于路由协议需要对链路进行双向检查，转发邻接将 LSP 链路发布给其他节点，此时需要再配置一条返回的 Tunnel，构成双向 Tunnel，并分别使能这两条隧道的转发邻接功能。

表 5-22　　配置转发捷径的步骤

步骤	命令	说明
1	**system-view** 例如：<Huawei> **system-view**	进入系统视图
2	**interface tunnel** *interface-number* 例如：[Huawei] **interface tunnel** 0/0/1	进入 MPLS TE 隧道的 Tunnel 接口视图
3	**mpls te igp shortcut** [**isis** | **ospf**] 例如：[Huawei-Tunnel0/0/1] **mpls te igp shortcut ospf**	配置 IGP Shortcut，使能 IGP 在进行增强的 SPF 计算（即进行 SPF 计算时包括了 TE 隧道）时使用处于 Up 状态的 MPLS TE 隧道功能。命令中的选项说明如下。 • **isis**：二选一可选项，指定使用 IS-IS 协议。 • **ospf**：二选一可选项，指定使用 OSPF 协议。 如果配置 IGP Shortcut 时不指定 IGP 类型，则缺省为 OSPF 和 IS-IS 都支持。 缺省情况下，IGP 在进行增强的 SPF 计算时使用 MPLS TE 隧道功能处于未使能状态，可用 **undo mpls te igp shortcut** 命令去使能 IGP 在进行增强的 SPF 计算时使用处于 Up 状态的 MPLS TE 隧道功能，使进行 SPF 计算时不包括 TE 隧道
4	**mpls te igp metric** { **absolute** *absolute-value* | **relative** *relative-value* } 例如：[Huawei-Tunnel0/0/1] **mpls te igp metric relative** -1	配置 TE 隧道的 IGP 度量值。因为 TE 隧道要作为逻辑链路 IGP 路由计算，所以肯定得指定这条逻辑链路对应的 Tunnel 接口的开销（度量）值，因为这是 IGP 路由计算的重要依据。命令中的参数说明如下。 • **absolute** *absolute-value*：二选一参数，指定绝对度量模式，整数形式，取值范围是 1～65535，TE 隧道的度量值就是配置的值。 • **relative** *relative-value*：二选一参数，指定相对度量模式，表示相对于隧道中物理链路的 IGP 度量差值，整数形式，取值范围是-10～10，缺省值为 0，TE 隧道的度量值是相应 IGP 路径度量值加上相对度量值。 要选择 TE 隧道，所配置的 TE 隧道度量值应小于其他可参与 IGP 路由计算的物理链路的开销值。 缺省情况下，TE 隧道使用的度量值与对应的物理链路的度量值相同，可用 **undo mpls te igp metric** 命令恢复缺省设置

（续表）

<table>
<tr><th>步骤</th><th>命令</th><th colspan="2">说明</th></tr>
<tr><td>5</td><td>mpls te commit
例如：[Huawei-Tunnel0/0/1] mpls te commit</td><td colspan="2">提交隧道当前配置，使配置生效</td></tr>
<tr><td rowspan="3">6</td><td>isis enable [process-id]
例如：[Huawei-Tunnel0/0/1] isis enable 1</td><td colspan="2">（二选一）使能隧道接口的 IS-IS Shortcut，当公网 TE 隧道配置使用 IS-IS 协议时采用。</td></tr>
<tr><td>ospf [process-id]
例如：[Huawei-Tunnel0/0/1] ospf 1</td><td rowspan="2">（二选一）使能隧道接口的 OSPF Shortcut，当公网 TE 隧道配置使用 OSPF 协议时采用</td><td>进入 OSPF 协议视图。</td></tr>
<tr><td>enable traffic-adjustment
例如：[Huawei-Tunnel0/0/1] enable traffic-adjustment</td><td>使能 OSPF Shortcut。缺省情况下，未使能 OSPF Shortcut，可用 undo enable traffic-adjustment 命令去使能 OSPF Shortcut 功能</td></tr>
</table>

表 5-23　　配置转发邻接的步骤

<table>
<tr><th>步骤</th><th>命令</th><th colspan="2">说明</th></tr>
<tr><td>1</td><td>system-view
例如：<Huawei> system-view</td><td colspan="2">进入系统视图</td></tr>
<tr><td>2</td><td>interface tunnel interface-number
例如：[Huawei] interface tunnel 0/0/1</td><td colspan="2">进入 MPLS TE 隧道的 Tunnel 接口视图。</td></tr>
<tr><td>3</td><td>mpls te igp advertise [hold-time interval]
例如：[Huawei-Tunnel0/0/1] mpls te igp advertise</td><td colspan="2">使能转发邻接将 MPLS TE 隧道作为虚拟链路发布到 IGP 网络的功能。可选参数 hold-time interval 用来指定 TE 隧道状态转为 Down 后到通知网络之前等待的时间，整数形式，取值范围是 0～4294967295，单位为毫秒。缺省值为 0。
缺省情况下，转发邻接将 MPLS TE 隧道作为虚拟链路发布到 IGP 网络的功能处于未使能状态，可用 undo mpls te igp advertise 命令去使能转发邻接将 MPLS TE 隧道作为虚拟链路发布到 IGP 网络的功能</td></tr>
<tr><td>4</td><td>mpls te igp metric { absolute absolute-value | relative relative-value }
例如：[Huawei-Tunnel0/0/1] mpls te igp metric relative -1</td><td colspan="2">配置 TE 隧道的 IGP 度量值。其他说明参见表 5-22 中的第 4 步</td></tr>
<tr><td>5</td><td>mpls te commit
例如：[Huawei-Tunnel0/0/1] mpls te commit</td><td colspan="2">提交隧道当前配置，使配置生效</td></tr>
<tr><td rowspan="3">6</td><td>isis enable [process-id]
例如：[Huawei-Tunnel0/0/1] isis enable 1</td><td colspan="2">（二选一）使能隧道接口的 IS-IS 转发邻接，当公网 TE 隧道配置使用 IS-IS 协议时采用</td></tr>
<tr><td>ospf [process-id]
例如：[Huawei-Tunnel0/0/1] ospf 1</td><td rowspan="2">（二选一）使能隧道接口的 OSPF 转发邻接，当公网 TE 隧道配置使用 OSPF 协议时采用</td><td>进入 OSPF 协议视图</td></tr>
<tr><td>enable traffic-adjustment advertise
例如：[Huawei-Tunnel0/0/1] enable traffic-adjustment advertise</td><td>使能 OSPF 转发邻接。缺省情况下，未使能 OSPF 转发邻接，可用 undo enable traffic-adjustment advertise 命令去使能 OSPF 转发邻接功能</td></tr>
</table>

已经完成将流量引入 MPLS TE 隧道的所有配置后，可通过以下 **display** 命令查看相

关配置，验证配置结果。

■ **display current-configuration**：查看将流量引入 MPLS TE 隧道的配置信息。

■ **display ip routing-table**：查看 MPLS TE 隧道接口作为路由出接口的情况。

■ **display ospf** [*process-id*] **traffic-adjustment**：查看与流量转发（转发捷径和转发邻接）相关的 OSPF 进程的隧道信息。

5.8.2 通过转发捷径将流量引入 TE 隧道的配置示例

转发捷径是一种将流量引入 TE 隧道的常用方式。在该方式中 TE 隧道将作为逻辑链路参与本地的 IGP 路由计算。通过设置 TE 隧道的 Metric 值，可以使 TE 隧道被优选，从而将流量引入 TE 隧道。

如图 5-22 所示，各设备之间通过 OSPF 协议实现路由互通，在 LSRA 上建立了一条经过 LSRB 到达 LSRC 的 TE 隧道；链路上标注的数字代表开销值。如果在 LSRA 上同时存在去往 LSRE 和 LSRC 的流量，则根据 OSPF 的选路结果，这两部分流量将都从接口 GE2/0/0 转发，因为这条路径中的开销最小。

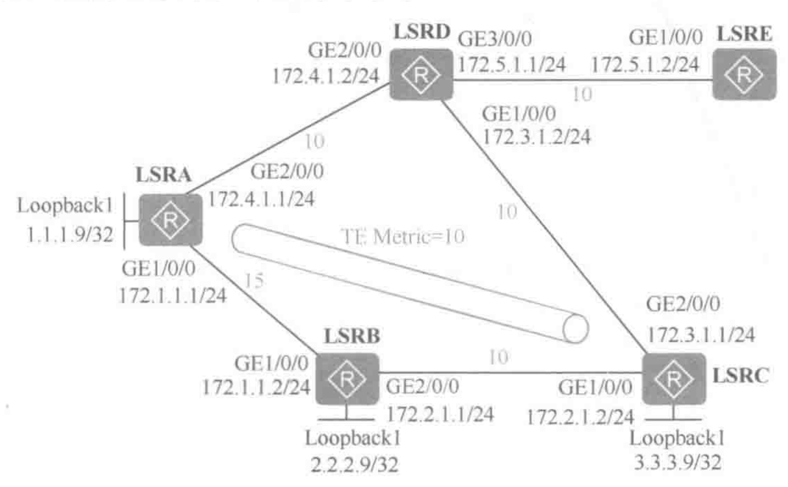


图 5-22 配置转发捷径将流量引入 TE 隧道示例的拓扑结构

现假设 LSRA 和 LSRD 之间链路的带宽为 100Mbit/s，去往 LSRC 的流量带宽需求为 50Mbit/s，去往 LSRE 的流量带宽需求为 60Mbit/s，两者相加为 110Mbit/s。此时，LSRA 和 LSRD 之间链路会发生拥塞，造成流量延迟或丢失。为了解决这个问题，可在 LSRA 上的 TE 隧道接口配置转发捷径，将去往 LSRC 的流量引入 TE 隧道。这样，这部分流量将从接口 GE1/0/0 转发，避免了网络拥塞的发生。

1. 基本配置思路分析

配置转发捷径将流量引入 TE 隧道的前提是要完成 MPLS TE 隧道的配置，所以本示例的基本配置任务中，首先是在 LSRA 到 LSRC 之间建立 MPLS TE 隧道的配置，然后再配置转发捷径功能，将 LSRA 到 LSRC 的流量引入到该 MPLS TE 隧道中。

根据 5.7 节及 5.8.1 节介绍的配置任务可得出本示例的基本配置思路如下。

（1）配置各设备接口（包括 Loopback 接口）的 IP 地址，并使用 OSPF 协议实现各节点之间路由可达，同时按图中标识配置各接口的 OSPF 开销值。

（2）在 LSRA、LSRB 和 LSRC 上配置 MPLS LSR ID，全局和公网接口上使能 MPLS、

MPLS TE、RSVP-TE 和 CSPF 能力，建立动态 MPLS TE 隧道。

（3）在 LSRA、LSRB 和 LSRC 上使能 OSPF TE，发布 TE 信息。

（4）在 LSRA 上创建并配置 Tunnel 接口目的 IP 地址等参数，配置显式路径。

（5）在 LSRA 的 TE 隧道接口下使能转发捷径功能，并配置 TE 隧道的 IGP 度量值。本示例中通过 LSRA 的 GE1/0/0 接口到达 LSRC 的链路的总开销值为 25，通过 GE2/0/0 到 LSRC 的链路的总开销为 20，要使 MPLS TE 隧道优选，只需把 Tunnel 接口的开销值配置小于 20 即可，因为 MPLS TE 隧道是点对点的，只需计算一个 Tunnel 接口的开销值即可。

2. 具体配置步骤

（1）配置各设备接口的 IP 地址和 OSPF 协议（所有设备都在 OSPF 1 进程，区域 0 中），并按图中标识配置各接口的 OSPF 开销值。

\# LSRA 上的配置。

```
<Huawei> system-view
[Huawei] sysname LSRA
[LSRA] interface gigabitethernet 1/0/0
[LSRA-GigabitEthernet1/0/0] ip address 172.1.1.1 255.255.255.0
[LSRA-GigabitEthernet1/0/0] ospf cost 15
[LSRA-GigabitEthernet1/0/0] quit
[LSRA] interface gigabitethernet 2/0/0
[LSRA-GigabitEthernet2/0/0] ip address 172.4.1.1 255.255.255.0
[LSRA-GigabitEthernet2/0/0] ospf cost 10
[LSRA-GigabitEthernet2/0/0] quit
[LSRA] interface loopback 1
[LSRA-LoopBack1] ip address 1.1.1.9 255.255.255.255
[LSRA-LoopBack1] quit
[LSRA] ospf 1
[LSRA-ospf-1] area 0
[LSRA-ospf-1-area-0.0.0.0] network 1.1.1.9 0.0.0.0
[LSRA-ospf-1-area-0.0.0.0] network 172.1.1.0 0.0.0.255
[LSRA-ospf-1-area-0.0.0.0] network 172.4.1.0 0.0.0.255
[LSRA-ospf-1-area-0.0.0.0] quit
[LSRA-ospf-1] quit
```

\# LSRB 上的配置。

```
<Huawei> system-view
[Huawei] sysname LSRB
[LSRB] interface gigabitethernet 1/0/0
[LSRB-GigabitEthernet1/0/0] ip address 172.1.1.2 255.255.255.0
[LSRB-GigabitEthernet1/0/0] ospf cost 15
[LSRB-GigabitEthernet1/0/0] quit
[LSRB] interface gigabitethernet 2/0/0
[LSRB-GigabitEthernet2/0/0] ip address 172.2.1.1 255.255.255.0
[LSRB-GigabitEthernet2/0/0] ospf cost 10
[LSRB-GigabitEthernet2/0/0] quit
[LSRB] interface loopback 1
[LSRB-LoopBack1] ip address 2.2.2.9 255.255.255.255
[LSRB-LoopBack1] quit
[LSRB] ospf 1
[LSRB-ospf-1] area 0
[LSRB-ospf-1-area-0.0.0.0] network 2.2.2.9 0.0.0.0
```

```
[LSRB-ospf-1-area-0.0.0.0] network 172.1.1.0 0.0.0.255
[LSRB-ospf-1-area-0.0.0.0] network 172.2.1.0 0.0.0.255
[LSRB-ospf-1-area-0.0.0.0] quit
[LSRB-ospf-1] quit
```

#　LSRC 上的配置。

```
<Huawei> system-view
[Huawei] sysname LSRC
[LSRC] interface gigabitethernet 1/0/0
[LSRC-GigabitEthernet1/0/0] ip address 172.2.1.2 255.255.255.0
[LSRC-GigabitEthernet1/0/0] ospf cost 10
[LSRC-GigabitEthernet1/0/0] quit
[LSRC] interface gigabitethernet 2/0/0
[LSRC-GigabitEthernet2/0/0] ip address 172.3.1.1 255.255.255.0
[LSRC-GigabitEthernet2/0/0] ospf cost 10
[LSRC-GigabitEthernet2/0/0] quit
[LSRC] interface loopback 1
[LSRC-LoopBack1] ip address 3.3.3.9 255.255.255.255
[LSRC-LoopBack1] quit
[LSRC] ospf 1
[LSRC-ospf-1] area 0
[LSRC-ospf-1-area-0.0.0.0] network 3.3.3.9 0.0.0.0
[LSRC-ospf-1-area-0.0.0.0] network 172.2.1.0 0.0.0.255
[LSRC-ospf-1-area-0.0.0.0] network 172.3.1.0 0.0.0.255
[LSRC-ospf-1-area-0.0.0.0] quit
[LSRC-ospf-1] quit
```

#　LSRD 上的配置。

```
<Huawei> system-view
[Huawei] sysname LSRD
[LSRD] interface gigabitethernet 1/0/0
[LSRD-GigabitEthernet1/0/0] ip address 172.3.1.2 255.255.255.0
[LSRD-GigabitEthernet1/0/0] ospf cost 10
[LSRD-GigabitEthernet1/0/0] quit
[LSRD] interface gigabitethernet 2/0/0
[LSRD-GigabitEthernet2/0/0] ip address 172.4.1.2 255.255.255.0
[LSRD-GigabitEthernet2/0/0] ospf cost 10
[LSRD-GigabitEthernet2/0/0] quit
[LSRD] interface gigabitethernet 3/0/0
[LSRD-GigabitEthernet3/0/0] ip address 172.5.1.2 255.255.255.0
[LSRD-GigabitEthernet3/0/0] ospf cost 10
[LSRD-GigabitEthernet3/0/0] quit
[LSRD] ospf 1
[LSRD-ospf-1] area 0
[LSRD-ospf-1-area-0.0.0.0] network 172.3.1.0 0.0.0.255
[LSRD-ospf-1-area-0.0.0.0] network 172.4.1.0 0.0.0.255
[LSRD-ospf-1-area-0.0.0.0] network 172.5.1.0 0.0.0.255
[LSRD-ospf-1-area-0.0.0.0] quit
[LSRD-ospf-1] quit
```

#　LSRE 上的配置。

```
<Huawei> system-view
[Huawei] sysname LSRE
[LSRE] interface gigabitethernet 1/0/0
[LSRE-GigabitEthernet1/0/0] ip address 172.5.1.2 255.255.255.0
[LSRE-GigabitEthernet1/0/0] ospf cost 10
[LSRE-GigabitEthernet1/0/0] quit
```

```
[LSRE] ospf 1
[LSRE-ospf-1] area 0
[LSRE-ospf-1-area-0.0.0.0] network 172.5.1.0 0.0.0.255
[LSRE-ospf-1-area-0.0.0.0] quit
[LSRE-ospf-1] quit
```

以上配置完成后，在 LSRA、LSRB、LSRC 节点上执行 **display ip routing-table** 命令，应可以看到相互之间都学到了到对方 Loopback1 的路由。

（2）在 LSRA、LSRB 和 LSRC 全局和公网接口上使能 MPLS、MPLS TE 和 RSVP-TE 能力，并在 LSRA 上全局使能 CSPF 能力。

LSRA 上的配置。

```
[LSRA] mpls lsr-id 1.1.1.9
[LSRA] mpls
[LSRA-mpls] mpls te
[LSRA-mpls] mpls rsvp-te
[LSRA-mpls] mpls te cspf
[LSRA-mpls] quit
[LSRA] interface gigabitethernet 1/0/0
[LSRA-GigabitEthernet1/0/0] mpls
[LSRA-GigabitEthernet1/0/0] mpls te
[LSRA-GigabitEthernet1/0/0] mpls rsvp-te
[LSRA-GigabitEthernet1/0/0] quit
```

LSRB 上的配置。

```
[LSRB] mpls lsr-id 2.2.2.9
[LSRB] mpls
[LSRB-mpls] mpls te
[LSRB-mpls] mpls rsvp-te
[LSRB-mpls] quit
[LSRB] interface gigabitethernet 1/0/0
[LSRB-GigabitEthernet1/0/0] mpls
[LSRB-GigabitEthernet1/0/0] mpls te
[LSRB-GigabitEthernet1/0/0] mpls rsvp-te
[LSRB-GigabitEthernet1/0/0] quit
[LSRB] interface gigabitethernet 2/0/0
[LSRB-GigabitEthernet2/0/0] mpls
[LSRB-GigabitEthernet2/0/0] mpls te
[LSRB-GigabitEthernet2/0/0] mpls rsvp-te
[LSRB-GigabitEthernet2/0/0] quit
```

LSRC 上的配置。

```
[LSRC] mpls lsr-id 3.3.3.9
[LSRC] mpls
[LSRC-mpls] mpls te
[LSRC-mpls] mpls rsvp-te
[LSRC-mpls] quit
[LSRC] interface gigabitethernet 1/0/0
[LSRC-GigabitEthernet1/0/0] mpls
[LSRC-GigabitEthernet1/0/0] mpls te
[LSRC-GigabitEthernet1/0/0] mpls rsvp-te
[LSRC-GigabitEthernet1/0/0] quit
```

（3）在 LSRA、LSRB 和 LSRC 上使能 OSPF TE，发布 TE 信息。

LSRA 上的配置。

```
[LSRA] ospf
[LSRA-ospf-1] opaque-capability enable  #---使能 OSPF 的 Opaque 能力
[LSRA-ospf-1] area 0
[LSRA-ospf-1-area-0.0.0.0] mpls-te enable   #---在当前 OSPF 区域使能 TE
[LSRA-ospf-1-area-0.0.0.0] quit
[LSRA-ospf-1] quit
```

LSRB 上的配置。

```
[LSRB] ospf
[LSRB-ospf-1] opaque-capability enable
[LSRB-ospf-1] area 0
[LSRB-ospf-1-area-0.0.0.0] mpls-te enable
[LSRB-ospf-1-area-0.0.0.0] quit
[LSRB-ospf-1] quit
```

LSRC 上的配置。

```
[LSRC] ospf
[LSRC-ospf-1] opaque-capability enable
[LSRC-ospf-1] area 0
[LSRC-ospf-1-area-0.0.0.0] mpls-te enable
[LSRC-ospf-1-area-0.0.0.0] quit
[LSRC-ospf-1] quit
```

（4）在 LSRA 上创建并配置 Tunnel 接口目的 IP 地址等参数，配置显式路径。

配置 TE 隧道的显式路径。

```
[LSRA] explicit-path pri-path   #---创建显式路径 pri-path
[LSRA-explicit-path-pri-path] next hop 172.1.1.2  #---配置严格下一跳 IP 地址 172.1.1.2
[LSRA-explicit-path-pri-path] next hop 172.2.1.2  #---配置严格下一跳 IP 地址 172.2.1.2
[LSRA-explicit-path-pri-path] next hop 3.3.3.9  #---配置严格下一跳 IP 地址 3.3.3.9
[LSRA-explicit-path-pri-path] quit
```

在 LSRA 上创建隧道接口。

```
[LSRA] interface tunnel 0/0/1
[LSRA-Tunnel0/0/1] ip address unnumbered interface loopback 1  #---配置 Tunnel0/0/1 接口借用 Loopback1 接口的 IP 地址
[LSRA-Tunnel0/0/1] tunnel-protocol mpls te   #---配置 Tunnel0/0/1 接口采用 TE 协议封装
[LSRA-Tunnel0/0/1] destination 3.3.3.9   #---配置 TE 隧道目的 IP 地址为 LSRC
[LSRA-Tunnel0/0/1] mpls te tunnel-id 100
[LSRA-Tunnel0/0/1] mpls te signal-protocol rsvp-te  #---配置隧道使用 RSVP-TE 作为信令协议
[LSRA-Tunnel0/0/1] mpls te path explicit-path pri-path  #---配置隧道应用的显式路径为 pri-path
[LSRA-Tunnel0/0/1] mpls te commit
[LSRA-Tunnel0/0/1] quit
```

（5）在 LSRA 的 TE 隧道接口下使能转发捷径功能，并配置 TE 隧道的 IGP 度量值为 10（绝对度量）。

```
[LSRA] interface tunnel 0/0/1
[LSRA-Tunnel0/0/1] mpls te igp shortcut ospf  #---配置使用 OSPF shortcut
[LSRA-Tunnel0/0/1] mpls te igp metric absolute 10  #---配置 TE 隧道绝对度量为 10
[LSRA-Tunnel0/0/1] mpls te commit
[LSRA-Tunnel0/0/1] quit
[LSRA] ospf 1
[LSRA-ospf-1] enable traffic-adjustment  #---使能转发捷径功能
[LSRA-ospf-1] quit
```

3. 配置结果验证

以上配置全部完成后，在 LSRA 上执行命令 **display ip routing-table** 3.3.3.9，可以看到去往 LSRC（3.3.3.9）的路由下一跳为 1.1.1.9，转发出接口为 Tunnel0/0/1，由此可证明去往 LSRC 的流量被引入到了 TE 隧道。

```
[LSRA] display ip routing-table 3.3.3.9
Route Flags: R - relay, D - download to fib
------------------------------------------------------------------------------
Routing Table : Public
Summary Count : 1
Destination/Mask    Proto   Pre  Cost      Flags NextHop         Interface

        3.3.3.9/32  OSPF    10   10          D   1.1.1.9         Tunnel0/0/1
```

5.8.3 通过转发邻接将流量引入 TE 隧道的配置示例

与转发捷径不同，配置转发邻接后，TE 隧道不但作为逻辑链路参与本地的 IGP 路由计算，还会被当作普通的 IGP 路由发布给邻居。通过设置 TE 隧道的 Metric 值，可以使包含 TE 隧道的路由被本地或其他设备优选，从而将流量引入 TE 隧道。

如图 5-23 所示，各设备之间通过 OSPF 协议实现路由互通，在 LSRA 上建立了一条经过 LSRB 到达 LSRC 的 TE 隧道；链路上标注的数字代表开销值。如果在 LSRA 和 LSRE 上同时存在去往 LSRC 的流量，则根据 OSPF 的选路结果，这两部分流量将都从 LSRD 的接口 GE1/0/0 转发。

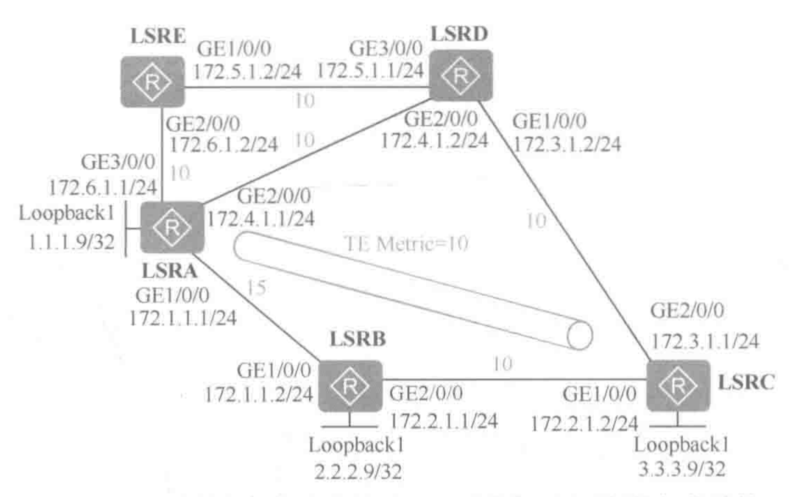


图 5-23 通过转发邻接将流量引入 TE 隧道配置示例的拓扑结构

现假设 LSRD 和 LSRC 之间链路的带宽为 100Mbit/s，LSRA 发往 LSRC 的流量带宽需求为 10Mbit/s，LSRE 发往 LSRC 的流量带宽需求为 100Mbit/s，两者相加为 110Mbit/s。此时，LSRD 和 LSRC 之间链路会发生拥塞，造成流量延迟或丢失。为了解决上述问题，可在 LSRA 上的 TE 隧道接口配置转发邻接。这样，从 LSRA 去往 LSRC 的流量将全部通过 TE 隧道转发；而从 LSRE 去往 LSRC 的流量一部分会通过 LSRD 转发，另一部分会发往 LSRA 并通过 TE 隧道转发，从而防止了 LSRD 和 LSRC 之间的链路发生拥塞。

说明 配置转发邻接后，LSRA 将把 TE 隧道作为 OSPF 路由发布给邻居。但由于 OSPF 协议需要对链路进行双向检查，因此还需要建立一条从 LSRC 到 LSRA 的 TE 隧道，并在隧道接口下使能转发邻接。这一点与转发捷径方式不一样，因为转发捷径方式中的 TE 隧道仅参与本地 IGP 路由计算，非 TE 隧道中的邻居设备不会感知隧道的存在的。

1. 基本配置思路分析

本示例的基本配置思路与上节介绍的配置示例的基本配置思路差不多，主要不同在于两个方面：（1）本示例采用的是通过转发邻接功能来引导流量进入 TE 隧道，（2）因为在配置采用转发邻接功能后，TE 隧道的入节点将会把 TE 隧道作为路由路径发给邻居设备，而邻居设备在进行 IGP 路由路径选择时需要进行双向检查，所以在本示例中要以隧道两端的 LSRA 和 LSRC 为入节点分别建立一条到达对端的 TE 隧道。

根据以上分析可得出本示例的基本配置思路如下。

（1）配置各设备接口（包括 Loopback 接口）的 IP 地址，并使用 OSPF 协议实现各节点之间路由可达，同时按图中标识配置各接口的 OSPF 开销值。

（2）在 LSRA、LSRB 和 LSRC 上配置 MPLS LSR ID，全局和公网接口上使能 MPLS、MPLS TE、RSVP-TE 和 CSPF 能力，建立动态 MPLS TE 隧道。

（3）在 LSRA、LSRB 和 LSRC 上使能 OSPF TE，发布 TE 信息。

（4）在 LSRA 和 LSRC 上分别创建并配置 Tunnel 接口目的 IP 地址等参数，配置显式路径。

（5）在 LSRA、LSRC 的 TE 隧道接口下分别使能转发邻接功能，并配置 TE 隧道的 IGP 度量值。这个 IGP 度量值不能随便设置，既要确保从 LSRA 到 LSRC 的流量能优先 TE 隧道，又要能使从 LSRE 到 LSRC 的流量有两条等价路由路径进行负载分担。通过计算，设置 TE 隧道的 IGP 度量值为 20 正好满足以上条件。

2. 具体配置步骤

（1）配置各接口的 IP 地址和 OSPF 协议（所有设备都在 OSPF 1 进程，区域 0 中），并按图中标识配置各接口的 OSPF 开销值。

LSRA 上的配置。

```
<Huawei> system-view
[Huawei] sysname LSRA
[LSRA] interface gigabitethernet 1/0/0
[LSRA-GigabitEthernet1/0/0] ip address 172.1.1.1 255.255.255.0
[LSRA-GigabitEthernet1/0/0] ospf cost 15
[LSRA-GigabitEthernet1/0/0] quit
[LSRA] interface gigabitethernet 2/0/0
[LSRA-GigabitEthernet2/0/0] ip address 172.4.1.1 255.255.255.0
[LSRA-GigabitEthernet2/0/0] ospf cost 10
[LSRA-GigabitEthernet2/0/0] quit
[LSRA] interface gigabitethernet 3/0/0
[LSRA-GigabitEthernet3/0/0] ip address 172.6.1.1 255.255.255.0
[LSRA-GigabitEthernet3/0/0] ospf cost 10
[LSRA-GigabitEthernet3/0/0] quit
[LSRA] interface loopback 1
[LSRA-LoopBack1] ip address 1.1.1.9 255.255.255.255
[LSRA-LoopBack1] quit
[LSRA] ospf 1
[LSRA-ospf-1] area 0
[LSRA-ospf-1-area-0.0.0.0] network 1.1.1.9 0.0.0.0
[LSRA-ospf-1-area-0.0.0.0] network 172.1.1.0 0.0.0.255
[LSRA-ospf-1-area-0.0.0.0] network 172.4.1.0 0.0.0.255
[LSRA-ospf-1-area-0.0.0.0] network 172.6.1.0 0.0.0.255
[LSRA-ospf-1-area-0.0.0.0] quit
[LSRA-ospf-1] quit
```

LSRB 上的配置。

```
<Huawei> system-view
[Huawei] sysname LSRB
[LSRB] interface gigabitethernet 1/0/0
[LSRB-GigabitEthernet1/0/0] ip address 172.1.1.2 255.255.255.0
[LSRB-GigabitEthernet1/0/0] ospf cost 15
[LSRB-GigabitEthernet1/0/0] quit
[LSRB] interface gigabitethernet 2/0/0
[LSRB-GigabitEthernet2/0/0] ip address 172.2.1.1 255.255.255.0
[LSRB-GigabitEthernet2/0/0] ospf cost 10
[LSRB-GigabitEthernet2/0/0] quit
[LSRB] interface loopback 1
[LSRB-LoopBack1] ip address 2.2.2.9 255.255.255.255
[LSRB-LoopBack1] quit
[LSRB] ospf 1
[LSRB-ospf-1] area 0
[LSRB-ospf-1-area-0.0.0.0] network 2.2.2.9 0.0.0.0
[LSRB-ospf-1-area-0.0.0.0] network 172.1.1.0 0.0.0.255
[LSRB-ospf-1-area-0.0.0.0] network 172.2.1.0 0.0.0.255
[LSRB-ospf-1-area-0.0.0.0] quit
[LSRB-ospf-1] quit
```

LSRC 上的配置。

```
<Huawei> system-view
[Huawei] sysname LSRC
[LSRC] interface gigabitethernet 1/0/0
[LSRC-GigabitEthernet1/0/0] ip address 172.2.1.2 255.255.255.0
[LSRC-GigabitEthernet1/0/0] ospf cost 10
[LSRC-GigabitEthernet1/0/0] quit
[LSRC] interface gigabitethernet 2/0/0
[LSRC-GigabitEthernet2/0/0] ip address 172.3.1.1 255.255.255.0
[LSRC-GigabitEthernet2/0/0] ospf cost 10
[LSRC-GigabitEthernet2/0/0] quit
[LSRC] interface loopback 1
[LSRC-LoopBack1] ip address 3.3.3.9 255.255.255.255
[LSRC-LoopBack1] quit
[LSRC] ospf 1
[LSRC-ospf-1] area 0
[LSRC-ospf-1-area-0.0.0.0] network 3.3.3.9 0.0.0.0
[LSRC-ospf-1-area-0.0.0.0] network 172.2.1.0 0.0.0.255
[LSRC-ospf-1-area-0.0.0.0] network 172.3.1.0 0.0.0.255
[LSRC-ospf-1-area-0.0.0.0] quit
[LSRC-ospf-1] quit
```

LSRD 上的配置。

```
<Huawei> system-view
[Huawei] sysname LSRD
[LSRD] interface gigabitethernet 1/0/0
[LSRD-GigabitEthernet1/0/0] ip address 172.3.1.2 255.255.255.0
[LSRD-GigabitEthernet1/0/0] ospf cost 10
[LSRD-GigabitEthernet1/0/0] quit
[LSRD] interface gigabitethernet 2/0/0
[LSRD-GigabitEthernet2/0/0] ip address 172.4.1.2 255.255.255.0
[LSRD-GigabitEthernet2/0/0] ospf cost 10
[LSRD-GigabitEthernet2/0/0] quit
[LSRD] interface gigabitethernet 3/0/0
```

```
[LSRD-GigabitEthernet3/0/0] ip address 172.5.1.1 255.255.255.0
[LSRD-GigabitEthernet3/0/0] ospf cost 10
[LSRD-GigabitEthernet3/0/0] quit
[LSRD] ospf 1
[LSRD-ospf-1] area 0
[LSRD-ospf-1-area-0.0.0.0] network 172.3.1.0 0.0.0.0
[LSRD-ospf-1-area-0.0.0.0] network 172.4.1.0 0.0.0.255
[LSRD-ospf-1-area-0.0.0.0] network 172.5.1.0 0.0.0.255
[LSRD-ospf-1-area-0.0.0.0] quit
[LSRD-ospf-1] quit
```

LSRE 上的配置。

```
<Huawei> system-view
[Huawei] sysname LSRE
[LSRE] interface gigabitethernet 1/0/0
[LSRE-GigabitEthernet1/0/0] ip address 172.5.1.2 255.255.255.0
[LSRE-GigabitEthernet1/0/0] ospf cost 10
[LSRE-GigabitEthernet1/0/0] quit
[LSRE] interface gigabitethernet 2/0/0
[LSRE-GigabitEthernet2/0/0] ip address 172.6.1.2 255.255.255.0
[LSRE-GigabitEthernet2/0/0] ospf cost 10
[LSRE-GigabitEthernet2/0/0] quit
[LSRE] ospf 1
[LSRE-ospf-1] area 0
[LSRE-ospf-1-area-0.0.0.0] network 172.5.1.0 0.0.0.0
[LSRE-ospf-1-area-0.0.0.0] network 172.6.1.0 0.0.0.255
[LSRE-ospf-1-area-0.0.0.0] quit
[LSRE-ospf-1] quit
```

以上配置完成后，在 LSRA、LSRB、LSRC 节点上执行 **display ip routing-table** 命令，应可以看到相互之间都学到了到对方 Loopback1 的路由。

（2）在 LSRA、LSRB 和 LSRC 上配置 MPLS LSR ID，在全局和公网接口上使能 MPLS TE、RSVP-TE，并在 LSRA 上全局使能 CSPF。

LSRA 上的配置。

```
[LSRA] mpls lsr-id 1.1.1.9
[LSRA] mpls
[LSRA-mpls] mpls te
[LSRA-mpls] mpls rsvp-te
[LSRA-mpls] mpls te cspf
[LSRA-mpls] quit
[LSRA] interface gigabitethernet 1/0/0
[LSRA-GigabitEthernet1/0/0] mpls
[LSRA-GigabitEthernet1/0/0] mpls te
[LSRA-GigabitEthernet1/0/0] mpls rsvp-te
[LSRA-GigabitEthernet1/0/0] quit
```

LSRB 上的配置。

```
[LSRB] mpls lsr-id 2.2.2.9
[LSRB] mpls
[LSRB-mpls] mpls te
[LSRB-mpls] mpls rsvp-te
[LSRB-mpls] quit
[LSRB] interface gigabitethernet 1/0/0
[LSRB-GigabitEthernet1/0/0] mpls
[LSRB-GigabitEthernet1/0/0] mpls te
```

```
[LSRB-GigabitEthernet1/0/0] mpls rsvp-te
[LSRB-GigabitEthernet1/0/0] quit
[LSRB] interface gigabitethernet 2/0/0
[LSRB-GigabitEthernet2/0/0] mpls
[LSRB-GigabitEthernet2/0/0] mpls te
[LSRB-GigabitEthernet2/0/0] mpls rsvp-te
[LSRB-GigabitEthernet2/0/0] quit
```

LSRC 上的配置。

```
[LSRC] mpls lsr-id 3.3.3.9
[LSRC] mpls
[LSRC-mpls] mpls te
[LSRC-mpls] mpls rsvp-te
[LSRC-mpls] quit
[LSRC] interface gigabitethernet 1/0/0
[LSRC-GigabitEthernet1/0/0] mpls
[LSRC-GigabitEthernet1/0/0] mpls te
[LSRC-GigabitEthernet1/0/0] mpls rsvp-te
[LSRC-GigabitEthernet1/0/0] quit
```

（3）在 LSRA、LSRB 和 LSRC 上使能 OSPF TE，发布 TE 信息。

LSRA 上的配置。

```
[LSRA] ospf
[LSRA-ospf-1] opaque-capability enable
[LSRA-ospf-1] area 0
[LSRA-ospf-1-area-0.0.0.0] mpls-te enable
[LSRA-ospf-1-area-0.0.0.0] quit
[LSRA-ospf-1] quit
```

LSRB 上的配置。

```
[LSRB] ospf
[LSRB-ospf-1] opaque-capability enable
[LSRB-ospf-1] area 0
[LSRB-ospf-1-area-0.0.0.0] mpls-te enable
[LSRB-ospf-1-area-0.0.0.0] quit
[LSRB-ospf-1] quit
```

LSRC 上的配置。

```
[LSRC] ospf
[LSRC-ospf-1] opaque-capability enable
[LSRC-ospf-1] area 0
[LSRC-ospf-1-area-0.0.0.0] mpls-te enable
[LSRC-ospf-1-area-0.0.0.0] quit
[LSRC-ospf-1] quit
```

（4）在 LSRA 和 LSRC 上分别创建 MPLS TE 隧道接口，配置并应用显式路径。

LSRA 上的配置。

```
[LSRA] explicit-path pri-path
[LSRA-explicit-path-pri-path] next hop 172.1.1.2
[LSRA-explicit-path-pri-path] next hop 172.2.1.2
[LSRA-explicit-path-pri-path] next hop 3.3.3.9
[LSRA-explicit-path-pri-path] quit
[LSRA] interface tunnel 0/0/1
[LSRA-Tunnel0/0/1] ip address unnumbered interface loopback 1
[LSRA-Tunnel0/0/1] tunnel-protocol mpls te
[LSRA-Tunnel0/0/1] destination 3.3.3.9
[LSRA-Tunnel0/0/1] mpls te tunnel-id 100
```

```
[LSRA-Tunnel0/0/1] mpls te path explicit-path pri-path
[LSRA-Tunnel0/0/1] mpls te commit
[LSRA-Tunnel0/0/1] quit
```

#　LSRC 上的配置。

```
[LSRC] explicit-path pri-path
[LSRC-explicit-path-pri-path] next hop 172.2.1.1
[LSRC-explicit-path-pri-path] next hop 172.1.1.1
[LSRC-explicit-path-pri-path] next hop 1.1.1.9
[LSRC-explicit-path-pri-path] quit
[LSRC] interface tunnel 0/0/1
[LSRC-Tunnel0/0/1] ip address unnumbered interface loopback 1
[LSRC-Tunnel0/0/1] tunnel-protocol mpls te
[LSRC-Tunnel0/0/1] destination 1.1.1.9
[LSRC-Tunnel0/0/1] mpls te tunnel-id 101
[LSRC-Tunnel0/0/1] mpls te path explicit-path pri-path
[LSRC-Tunnel0/0/1] mpls te commit
[LSRC-Tunnel0/0/1] quit
```

（5）在 LSRA 和 LSRC 的 TE 隧道接口下使能转发邻接，配置 IGP 度量值为 10（绝对度量），使 TE 隧道对应的逻辑链路在 IGP 路由计算中被优选，同时使 LSRE 去往 LSRC 的流量有两条等价路由路径，分别是通过 LSRD 到达和通过 TE 隧道到达，总路径开销均为 20。

#　LSRA 上的配置。

```
[LSRA] interface tunnel 0/0/1
[LSRA-Tunnel0/0/1] mpls te igp advertise
[LSRA-Tunnel0/0/1] mpls te igp metric absolute 10
[LSRA-Tunnel0/0/1] mpls te commit
[LSRA-Tunnel0/0/1] quit
[LSRA] ospf 1
[LSRA-ospf-1] enable traffic-adjustment advertise
[LSRA-ospf-1] quit
```

#　LSRC 上的配置。

```
[LSRC] interface tunnel 0/0/1
[LSRC-Tunnel0/0/1] mpls te igp advertise
[LSRC-Tunnel0/0/1] mpls te igp metric absolute 10
[LSRC-Tunnel0/0/1] mpls te commit
[LSRC-Tunnel0/0/1] quit
[LSRC] ospf 1
[LSRC-ospf-1] enable traffic-adjustment advertise
[LSRC-ospf-1] quit
```

3．配置结果验证

以上配置全部完成后，在 LSRA 上执行命令 **display ip routing-table** 3.3.3.9，可以看到去往 LSRC（3.3.3.9）的路由下一跳为 1.1.1.9，转发出接口为 Tunnel0/0/1，去往 LSRC 的流量被引入了 TE 隧道。

```
[LSRA] display ip routing-table 3.3.3.9
Route Flags: R - relay, D - download to fib
------------------------------------------------------------------------------
Routing Table : Public
Summary Count : 1
Destination/Mask    Proto   Pre  Cost      Flags NextHop         Interface

        3.3.3.9/32  OSPF    10   10          D   1.1.1.9         Tunnel0/0/1
```

在 LSRE 上执行命令 **display ip routing-table** 3.3.3.9，可以看到去往 LSRC（3.3.3.9）有两条等价路由（开销值均为 20）。去往 LSRC 的流量将会有一部分通过 LSRD 转发，另一部分会发往 LSRA 并通过 TE 隧道转发。

```
[LSRE] display ip routing-table 3.3.3.9
Route Flags: R - relay, D - download to fib
------------------------------------------------------------------------------
Routing Table : Public
Summary Count : 2
Destination/Mask    Proto   Pre  Cost      Flags NextHop         Interface

        3.3.3.9/32  OSPF    10   20           D  172.5.1.1       GigabitEthernet1/0/0
                    OSPF    10   20           D  172.6.1.1       GigabitEthernet2/0/0
```

5.9 MPLS TE 隧道维护

MPLS TE 的维护包括：检测 TE 隧道的连通性、查看隧道错误信息、清除运行信息、重启 RSVP 进程等。这在日常的 MPLS TE 隧道维护中可能经常要用到，以查看相关配置、排除隧道故障。

1. 检测 TE 隧道的连通性

■ **ping lsp** [**-a** *source-ip* | **-c** *count* | **-exp** *exp-value* | **-h** *ttl-value* | **-m** *interval* | **-r** *reply-mode* | **-s** *packet-size* | **-t** *time-out* | **-v**] * **te tunnel** *interface-number* [**hot-standby**]：检测 TE 隧道从入节点到出节点是否连通。如果指定 **hot-standby** 参数，则可以实现对备份 CR-LSP 的检测。

■ **tracert lsp** [**-a** *source-ip* | **-exp** *exp-value* | **-h** *ttl-value* | **-r** *reply-mode* | **-t** *time-out*] * **te tunnel** *interface-number* [**hot-standby**]：查看数据包从 TE tunnel 入节点到出节点所经过的网关。如果指定 **hot-standby** 参数，则可以实现对备份 CR-LSP 的检测。

2. 配置 MPLS TE 告警上报功能

为了方便运维，及时了解 MPLS 网络的运行状态，可以配置 MPLS TE 告警上报功能，将 RSVP 和 MPLS TE 隧道的状态变化和动态标签的使用情况通知给网管系统，提醒用户注意。

■ 配置 RSVP 的 Trap 功能

在系统视图下执行 **snmp-agent trap enable feature-name mpls_rsvp** [**trap-name** *trap-name*]命令，打开 MPLS RSVP 模块的告警开关。参数 **trap-name** *trap-name* 用来打开或关闭指定名称告警的开关，具体参见对应产品手册说明。如果不指定本参数，则打开或关闭 MPLS RSVP 模块所有告警的开关。缺省情况下，MPLS RSVP 模块的告警开关处于关闭状态。

可通过 **display snmp-agent trap feature-name mpls_rsvp all** 命令查看 MPLS RSVP 模块的所有告警开关的状态信息。

■ 配置 TE 隧道的 Trap 功能

在系统视图下执行 **snmp-agent trap enable feature-name tunnel-te** [**trap-name** *trap-*

name]命令打开 Tunnel-TE 模块的告警开关。参数 **trap-name** *trap-name* 用来打开或关闭指定名称告警的开关，具体参见对应产品手册说明。如果不指定本参数，则打开或关闭 Tunnel-TE 模块所有告警的开关。缺省情况下，Tunnel-TE 模块的告警开关处于关闭状态。

可通过 **display snmp-agent trap feature-name tunnel-te all** 命令查看 Tunnel-TE 模块的所有告警开关的状态信息。

3. 清除运行信息

■ **reset mpls rsvp-te statistics** { **global** | **interface** [*interface-type interface-number*] }：清除 RSVP-TE 的运行统计信息。

■ **reset mpls stale-interface** [*interface-index*]：清除处于 stale 状态的 MPLS 接口的信息。

4. 查看 TE 信息

■ **display default-parameter mpls te management**：查看 MPLS TE 管理默认配置参数值。

■ **display mpls te tunnel statistics** 或者 **display mpls lsp statistics**：查看隧道统计信息。

■ **display mpls te tunnel-interface last-error** [*tunnel-name*]：查看 Tunnel 接口的错误信息。

■ **display mpls te tunnel-interface failed**：查看未建立成功或者正在建立的 MPLS TE 隧道。

■ **display mpls te tunnel-interface traffic-state** [*tunnel-name*]：查看本地节点的隧道接口当前的流量状态。

■ **display mpls rsvp-te statistics** { **global** | **interface** [*interface-type interface- number*] }：查看 RSVP-TE 运行统计信息。

5. 重启隧道接口

需要使隧道的相关配置立即生效时，可以在 Tunnel 接口视图下配置 **mpls te commit** 命令，并在用户视图下执行下面的 **reset mpls te tunnel-interface tunnel** *interface-number* 命令重启隧道接口。

6. 重启 RSVP 进程

当需要重建所有的 RSVP 类型的 CR-LSP 或验证 RSVP 工作过程时，可在用户视图下执行 **reset mpls rsvp-te** 命令重启 RSVP 进程。

第6章 MPLS TE参数调整配置与管理

6.1 调整RSVP-TE信令参数

6.2 调整CR-LSP的路径选择

6.3 调整MPLS TE隧道的建立



第 5 章介绍了 MPLS TE 隧道基本功能的配置与管理方法，本章具体介绍 MPLS TE 隧道在参数调整方面的配置与管理方法。

在参数调整方面主要包括 RSVP 协议发送消息的参数或参数格式、CR-LSP 路径选择参数（如路径仲裁方式、路径度量、链路管理组、亲和属性等）、RSVP 邻居认证功能的配置，这对优化 CR-LSP 路径选择和隧道通信性能非常有用，可根据实际应用需求选择其中一项或多项参数进行调整。

6.1 调整 RSVP-TE 信令参数

RSVP-TE 是 MPLS TE 和第 9 章介绍的 MPLS DS-TE 隧道建立的信令协议。在完成动态 MPLS TE，或将动态 DS-TE 隧道基本功能配置后，还可通过调整 RSVP-TE 信令参数，进一步满足用户对可靠性和网络资源充分利用的需求。

RSVP 信令参数涉及资源预留风格、预留确认、状态定时器、摘要刷新、Hello 扩展、邻居认证等方面，通过配置相应的参数可以灵活的实现用户需求。当然这些参数调整都是可选配置的，且无配置顺序要求，可根据实际需要选择配置。

6.1.1 配置 RSVP 资源预留风格

资源预留风格是当中间节点设备收到入节点的资源预留请求时，决定是要为某 CR-LSP 分配独占的预留资源，还是与其他 CR-LSP 共享预留资源。很显然，这主要针对有多条 CR-LSP 经过该中间节点时才需要考虑的，但资源预留风格需要在 MPLS 隧道入节点上配置，使得有多条 CR-LSP 经过同一中间节点、在处理资源预留请求时采用在入节点上配置资源预留方式。

在第 5 章 5.1.2 节已介绍，目前华为设备仅支持 FF（Fixed Filter，固定过滤风格）和 SE（Shared Explicit，共享显式风格）两种资源预留风格。FF 风格是中间节点为每个资源预留请求的发送者（入节点）创建单独的预留，该预留不与其他发送者共享。SE 风格是为所有发送者预留单一的资源，该预留允许指定的一系列发送者共享。

RSVP 资源预留风格需在 MPLS TE 隧道**入节点**进行表 6-1 所示配置，中间节点按照入节点的配置采用对应的预留风格。

表 6-1 RSVP 资源预留风格的配置步骤

步骤	命令	说明
1	**system-view** 例如：<Huawei> **system-view**	进入系统视图
2	**interface tunnel** *tunnel-number* 例如：[Huawei] **interface tunnel** 0/0/1	进入 MPLS TE 隧道的 Tunnel 接口视图
3	**mpls te resv-style** { **ff** | **se** } 例如：[Huawei-Tunnel0/0/1] **mpls te resv-style ff**	配置隧道的资源预留风格。命令中的选项说明如下。 • **ff**：二选一选项，指定使用固定过滤器类型 FF（Fixed-Filter）样式的资源预留。 • **se**：二选一选项，指定使用共享显式类型 SE（Shared-Explicit）样式的资源预留

（续表）

步骤	命令	说明
3	**mpls te resv-style** { **ff** | **se** } 例如：[Huawei-Tunnel0/0/1] **mpls te resv-style ff**	缺省情况下，资源预留风格为共享显式类型 SE，可用 **undo mpls te resv-style** 命令恢复缺省的资源预留样式
4	**mpls te commit** 例如：[Huawei-Tunnel0/0/1] **mpls te commit**	提交隧道配置，使以上配置更改生效。凡发生了 TE 隧道配置更改，均要执行本命令，才能使配置更改生效

6.1.2 配置 RSVP-TE 预留确认

当入节点在收到出节点通过 RSVP-TE 信令协议发来的 Resv 消息后，如果 Resv 消息中没有包括 RESV_CONFIRM 的对象，则在通过 ResvConf 消息向出节点进行确认时只代表入节点收到了 Resv 消息，并且在入节点处成功保留了相应资源，并不代表已成功为对应 CR-LSP 建立了资源预留，因为这些资源接下来仍可能被其他应用抢占。

此时可通过在出节点上启动预留确认机制，这样就会在向入节点方向逐跳发送的 Resv 消息中包括一个名为 RESV_CONFIRM 的对象，要求沿途各节点为对应 CR-LSP 进行资源预留，最后入节点向出节点发送的 ResvConf 消息中也会对预留进行确认。

配置 RSVP-TE 预留确认的方法是在 MPLS TE 隧道**出节点**的 MPLS 视图下执行 **mpls rsvp-te resvconfirm** 命令，使能节点的预留确认机制。缺省情况下，节点上的预留确认机制处于未使能状态，可用 **undo mpls rsvp-te resvconfirm** 命令去使能节点的预留确认机制。

6.1.3 配置 RSVP 的状态定时器

动态建立的 CR-LSP 要依靠 RSVP Refresh 消息来维护。CR-LSP 建立过程中的资源预留状态包括路径状态（Path State）和预留状态（Reservation State）。这两种状态分别由 Path 消息和 Resv 消息创建并定时刷新，将定时刷新时发送的 Path 消息和 Resv 消息统称为 RSVP Refresh 消息。通过 RSVP 消息的定时刷新来维持节点上资源预留状态的机制称为 RSVP-TE 的“软状态”。

RSVP Refresh 消息用于在 RSVP 邻居节点进行状态同步，消息内容分别包含路径状态块（PSB）和预留状态块（RSB）。对于路径状态或者预留状态，如果连续一段时间没有收到刷新消息，这个状态将被删除。RSVP Refresh 消息除了可以进行节点间状态同步之外，还可以检测各邻居间的可达性，维护 RSVP 节点之间的邻居关系。

RSVP 消息是以 IP 数据报的形式传输的，因此 RSVP 消息的传输是不可靠的。在 CR-LSP 建立后，通过软状态机制同步 RSVP 邻居节点的状态（包括 PSB 和 RSB），各节点仍然会周期性地向上下游邻居节点发送 RSVP Refresh 消息。但 Refresh 消息并不是一种新的消息，而是以前发布过的 Path 或 Resv 消息的再次传送。所以，要适度控制 RSVP Refresh 消息的发送频率，否则会严重占用隧道的带宽资源。这个刷新周期就是 Path 消息或 Resv 消息中的 TIME_VALUE 对象中指定的时间间隔。

通过配置 RSVP 的状态定时器，可以设置 RSVP 刷新消息的发送时间间隔和重发次数，从而改变这个超时时间，建议使用缺省配置。通过以下公式计算超时时间。

超时时间 ＝ （*keep-multiplier-number*＋0.5） × 1.5 × *refresh-interval*

其中，*keep-multiplier-number* 代表 RSVP 刷新消息重发的次数，*refresh-interval* 代表 RSVP 刷新消息的发送时间间隔。

如果要修改缺省的 RSVP 状态定时器，则需要在路径的**各节点**上做相同的配置，具体配置方法见表 6-2。

表 6-2　　RSVP 状态定时器的配置步骤

步骤	命令	说明
1	**system-view** 例如：<Huawei> **system-view**	进入系统视图
2	**mpls** 例如：[Huawei] **mpls**	进入 MPLS 视图
3	**mpls rsvp-te timer refresh** *refresh-interval* 例如：[Huawei-mpls] **mpls rsvp-te timer refresh** 60	设置节点的 RSVP 刷新消息的发送时间间隔，整数形式，取值范围是 10～65535，单位是秒。修改后需要等到上次定时器超时以后才生效。建议不要设置超长刷新周期或反复修改刷新周期。 缺省情况下，节点的 RSVP 刷新消息的发送时间间隔为 30s，可用 **undo mpls rsvp-te timer refresh** 命令恢复缺省设置
4	**mpls rsvp-te keep-multiplier** *keep-multiplier-number* 例如：[Huawei-mpls] **mpls rsvp-te keep-multiplier** 5	设置 RSVP 刷新消息重发的次数。参数 *keep-multiplier-number* 为整数形式，取值范围是 3～255。这个重发次数是指在一个超时定时器时间内可以重发 Path 消息或 Resv 消息的次数。 超时定时器＝（*keep-multiplier-number*＋0.5） × 1.5 × *refresh-interval* 缺省情况下，RSVP 刷新消息重发的次数为 3，可用 **undo mpls rsvp-te keep-multiplier** 命令恢复缺省设置

6.1.4 使能 RSVP-TE 摘要刷新功能

通过 6.1.3 节介绍的 RSVP 状态定时器的调整可在一定程度上减轻频繁发送重复的 RSVP 刷新消息对隧道带宽占用的影响，但是这种“软状态”机制所采用的 Path 消息和 Resv 消息的报文长度较大，当建立的 CR-LSP 很多时仍会过多的占用隧道带宽资源。为此可通过配置 RSVP 摘要刷新功能解决这个问题。

RSVP 摘要刷新是通过在原有 RSVP 协议中定义新的对象来实现的，具体包括以下两个方面。

（1）Message_ID 扩展和重传机制

RFC2961 中定义的 Message_ID 扩展机制是在 RSVP 消息中携带扩展的对象。其中，Message_ID 和 Message_ID_ACK 对象用于 RSVP 消息确认，从而提高 RSVP 消息的可靠性。

使用 Message_ID 扩展对象还可实现 RSVP 重传机制。节点发送携带 Message_ID 的 RSVP 消息后初始化重传时间（假设为 *Rf* 秒）。如果在 Rf 时间间隔内没有收到 ACK 消息，经过（1＋*Delta*）×*Rf* 后，将重传此消息。*Delta* 取决于发送方增加重传间隔的速率。重传将一直持续，直到收到一个确认消息或重传次数达到允许的最大限制值（称为重传增量）。

（2）摘要刷新 Srefresh（Summary Refresh）

摘要刷新不传送标准的 Path 或 Resv 消息，而仍能实现对 RSVP 状态的刷新。使用摘要刷新的好处是它减少了维持 RSVP 状态所需传输及处理的信息量。使用专门的 Srefresh 消息来更新 RSVP 状态，常规的刷新消息将被抑制。

摘要刷新需要与 Message_ID 扩展配合使用，因为摘要刷新消息承载了一系列 Message_ID 对象，用于识别需要被刷新的 Path 及 Resv 状态。只有那些已经包含 Message_ID 的 Path 和 Resv 消息发布过的状态才能使用摘要刷新机制刷新。

当节点接收到一条摘要刷新消息时，与本地状态块（PSB 或 RSB）进行匹配。如果匹配，就更新本地状态，就像接收到一个标准的 RSVP 刷新消息一样；如果不匹配，节点将发送一个 NACK 消息来通知摘要刷新消息的发送者，并根据 Path 或 Resv 消息刷新相应的 PSB 或 RSB，同时更新 Message_ID。

Message_ID 对象中包含 Message_ID 序列号。当 CR-LSP 发生变化时，相应的 Message_ID 序列号增大。这样一来，在节点收到 Path 消息时，将其中的 Message_ID 序列号与本地状态块中保存的 Message_ID 序列号比较，如果序列号相等，则保持状态不变；如果前者大于后者，则表示状态已更新。

可在两个邻居节点的接口视图下或 MPLS 视图下使能全局的摘要刷新（Srefresh）功能，以减少刷新消息而带来的额外开销，提高网络性能。在接口视图下使能摘要刷新功能后，则当前接口会使用 Srefresh 消息替代 Path、Resv 消息来刷新 RSVP 软状态。在 MPLS 视图下使能摘要刷新功能后，则节点上所有接口都使用 Srefresh 消息替代 Path、Resv 消息来刷新 RSVP 软状态，具体配置方法见表 6-3，可在 MPLS TE 隧道各节点上配置，建议但不强制每个节点的每个接口都同步使能。

表 6-3　　使能 **RSVP-TE** 摘要刷新功能的步骤

步骤	命令	说明
1	**system-view** 例如：<Huawei> **system-view**	进入系统视图
方法一：在 MPLS 视图下配置 RSVP-TE 摘要刷新功能		
2	**mpls** 例如：[Huawei] **mpls**	进入 MPLS 视图
3	**mpls rsvp-te srefresh** 例如：[Huawei-mpls] **mpls rsvp-te srefresh**	使能全局的摘要刷新功能。如果接口上没有单独使能 RSVP 摘要刷新功能，则继续全局下的配置，否则以对应接口下的配置为准。 缺省情况下，未使能全局的摘要刷新功能，可用 **undo mpls rsvp-te srefresh** 命令去使能全局的摘要刷新功能
方法二：在接口视图下配置 RSVP-TE 摘要刷新功能		
2	**interface** *interface-type interface-number* 例如：[Huawei] **interface** gigabitethernet 1/0/0	进入 MPLS TE 链路的接口视图
3	**mpls rsvp-te srefresh** 例如：[Huawei-GigabitEthernet1/0/0] **mpls rsvp-te srefresh**	使能以上接口的摘要刷新功能。接口上的配置优先级高于在 MPLS 视图下的全局配置。 缺省情况下，未使能接口的摘要刷新功能，可用 **undo mpls rsvp-te srefresh** 命令去使能接口的摘要刷新功能

（续表）

步骤	命令	说明
4	**mpls rsvp-te timer retransmission** { **increment-value** *increment* | **retransmit-value** *interval* } * 例如：[Huawei-GigabitEthernet1/0/0] **mpls rsvp-te timer retransmission retransmit-value** 500 **increment-value** 2	（可选）配置重传参数。命令中的参数说明如下。 ● **increment-value** *increment*：可多选参数，指定重传增量值，整数形式，取值范围是 1～10。达到这个值后，在本次重传定时器时间内，对应的 RSVP 消息不能再次重传。 ● **retransmit-value** *interval*：可多选参数，指定 RSVP 消息发送后没有收到对应的 ACK 消息时，下次进行重传时所需等待的时间间隔，即重传间隔，整数形式，取值范围是 500～5000，单位是毫秒。 【注意】重传间隔和重传增量之间的关系如下。 下一次的重传间隔 *interval* = 当前 *interval* x (1 + *increment*)。 缺省情况下，重传增量值为 1，重传定时器间隔为 5000ms，可用 **undo mpls rsvp-te timer retransmission** [**increment-value** [*increment*] | **retransmit-value** [*interval*]] * 命令恢复对应参数为缺省配置

6.1.5 配置 RSVP 的 Hello 扩展

RSVP Refresh 消息除了可以进行节点间 PSB 和 RSB 状态同步之外，另外还可以检测各邻居间的可达性，维护 RSVP 节点之间的邻居关系。但是这种软状态机制是采用 Path 消息和 Resv 消息进行的，检测速度较慢，在路径出现故障时不能及时触发业务向备份路径切换流量。因此，引入 RSVP Hello 扩展（是原来 RSVP Hello 机制针对 TE 隧道的扩展）来解决这个问题。RSVP Hello 适用于 TE FRR（快速重路由）和 RSVP GR（平滑重启）的场景中。

RSVP 的 Hello 扩展机制用于快速检测 RSVP 邻居节点的可达性，当检测到 RSVP 邻居节点不可达时，相关的 MPLS TE 隧道将被拆除。同时还可以检测邻居节点是否处于重启状态，以支持邻居实现 RSVP GR。

下面以图 6-1 所示的示例介绍 RSVP Hello 机制的实现过程。LSRA、LSRB 之间有链路直接相连。

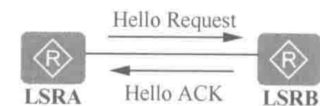


图 6-1 Hello 握手机制示例

（1）当 LSRA 接口下使能了 RSVP Hello 时，LSRA 会向 LSRB 发送 Hello Request 消息。

（2）若 LSRB 收到了 Hello 消息，并且 LSRB 也使能了 RSVP Hello，就会给 LSRA 节点回复 Hello ACK 消息。

（3）LSRA 收到 LSRB 的 Hello ACK 消息后，就确认 LSRA 的邻居 LSRB 是可达的。当 LSRA 连续三次向 LSRB 发送 Hello Request 消息后，LSRB 仍然没有给 LSRA 回 Hello ACK 消息，此时就认为 LSRB 邻居丢失，触发 TE FRR 切换并重新初始化 RSVP Hello。

RSVP 的 Hello 扩展机制还用于检测邻居重启。在图 6-1 中，当 LSRA 和 LSRB 都使能 RSVP GR 功能时，在 LSRA 检查到邻居 LSRB 丢失后，等待 LSRB 发送有 GR 扩展的 Hello Request 消息。收到此消息后，LSRA 开始协助 LSRB 恢复 RSVP 状态，并向 LSRB 发送 Hello ACK 消息。LSRB 收到 LSRA 回复的 Hello ACK 消息后，发现 LSRA 已开始协助自己进行 GR，然后 LSRA 和 LSRB 互通 Hello 消息，维持 GR 恢复状态。

配置 RSVP 的 Hello 扩展的步骤见表 6-4，需分别在 MPLS TE 隧道各节点的全局和公网接口上同时配置。

表 6-4　**RSVP 的 Hello 扩展的配置步骤**

步骤	命令	说明
1	**system-view** 例如：<Huawei> **system-view**	进入系统视图
2	**mpls** 例如：[Huawei] **mpls**	进入 MPLS 视图
3	**mpls rsvp-te hello** 例如：[Huawei-mpls] **mpls rsvp-te hello**	使能本节点的 RSVP Hello 扩展功能。 缺省情况下，未使能全局 RSVP 的 Hello 扩展功能，可用 **undo mpls rsvp-te hello** 命令去使能全局的 RSVP Hello 扩展功能
4	**mpls rsvp-te hello-lost** *times* 例如：[Huawei-mpls] **mpls rsvp-te hello-lost** 5	配置允许 Hello 消息丢失的最大次数，整数形式，取值范围是 3～10。 缺省情况下，启用 Hello 扩展机制后，连续未收到 3 次 Hello 应答消息即认为节点发生故障，相关的 MPLS TE 隧道将被拆除，可用 **undo mpls rsvp-te hello-lost** 命令恢复缺省值
5	**mpls rsvp-te timer hello** *interval* 例如：[Huawei-mpls] **mpls rsvp-te timer hello** 10	配置 Hello 消息刷新时间间隔，整数形式，取值范围是 1～25，单位是秒。新修改的刷新周期配置要等到上次定时器超时以后才生效。 启用 Hello 扩展机制后，缺省的 Hello 消息刷新时间间隔为 3s，可用 **undo mpls rsvp-te timer hello** 命令恢复缺省设置
6	**quit** 例如：[Huawei-mpls] **quit**	返回系统视图
7	**interface** *interface-type interface-number* 例如：[Huawei] **interface** gigabitethernet 1/0/0	进入 MPLS TE 链路的接口视图
8	**mpls rsvp-te hello** 例如：[Huawei-GigabitEthernet1/0/0] **mpls rsvp-te hello**	使能以上接口的 RSVP Hello 扩展功能 缺省情况下，未使能接口 RSVP 的 Hello 扩展功能，可用 **undo mpls rsvp-te hello** 命令去使能接口的 RSVP Hello 扩展功能

6.1.6　配置 RSVP 消息格式

RSVP 消息格式中包括了许多可选对象，而不同厂商中所采用的可选对象或对象的编码格式不完全一样，这样就造成了不同厂商设备的 RSVP 消息格式可能不完全相同。这时就可以调整华为设备的 RSVP 消息格式，以便与其他厂商设备互通。

可在 MPLS TE 隧道**各节点上**按表 6-5 的步骤配置 RSVP 发送消息携带的对象格式。可在 MPLS TE 隧道**中间节点和出节点上**按表 6-6 的步骤调整华为设备 Resv 消息携带的 RRO 格式，与其他厂商设备保持一致，以实现顺利互通。

表 6-5 配置 **RSVP** 发送消息携带的对象格式的步骤

步骤	命令	说明
1	**system-view** 例如：<Huawei> **system-view**	进入系统视图
2	**mpls** 例如：[Huawei] **mpls**	进入 MPLS 视图
3	**mpls rsvp-te send-message { suggest-label | extend-class-type value-length-type | session-attribute without-affinity | down-reason }** 例如：[Huawei-mpls] **mpls rsvp-te send-message session-attribute without-affinity**	配置 RSVP 发送消息携带的对象格式。命令中的选项说明如下。 • **suggest-label**：多选一选项，指定 RSVP 消息中携带 suggest-label 对象。在 GR（平滑启动）结束后，当上游为华为设备，下游为其他厂商设备时，上游设备向下游设备发送 Path 消息中默认不携带 suggest-label 对象。但如果其他厂商设备要求携带 suggest-label，则需要指定本选项。 • **extend-class-type value-length-type**：多选一选项，指定 RSVP 消息中携带的 extend-class-type 对象的编码格式是 value-length-type（即 VTL 格式）。华为设备 extend-calss-type 对象的编码格式为 TLV，如果其他厂商设备该对象的编码格式为 VLT，则需要选择本选项。 • **session-attribute without-affinity**：多选一选项，指定 RSVP 消息中的 session-attribute 对象不携带 affinity（亲和）属性。如果上游为华为设备，支持 session-attribute 对象，下游为其他厂商设备，不支持 session-attribute 对象，且上游的华为设备在配置 TE 隧道时配置了亲和属性，则需要选择本选项。 • **down-reason**：多选一选项，指定 RSVP 消息中携带 down-reason 对象。如果想在隧道入节点查看中间节点和出节点记录的隧道故障原因，则要在**中间节点和出节点**配置 **mpls rsvp-te send-message down-reason** 命令。 该命令涉及发送消息携带的四种对象格式配置后可以同时生效。但修改配置时，只能对新创建的 LSP 生效。 缺省情况下，没有配置发送消息中携带的对象格式，可用 **undo mpls rsvp-te send-message { suggest-label | extend-class-type value-length-type | session-attribute without-affinity | down-reason }** 命令恢复对应对象的缺省配置

表 6-6 配置 **Resv** 消息携带的记录路由对象 **RRO** 格式的步骤

步骤	命令	说明
1	**system-view** 例如：<Huawei> **system-view**	进入系统视图
2	**mpls** 例如：[Huawei] **mpls**	进入 MPLS 视图
3	**mpls rsvp-te resv-rro transit { { incoming | incoming-with-label } | { routerid | routerid-with-label } | { outgoing | outgoing-with-label } }** * 例如：[Huawei-mpls] **mpls rsvp-te resv-rro transit routerid-with-label incoming outgoing-with-label**	（二选一）在 **Transit 节点**上配置 Resv 消息携带的 RRO 对象格式。命令中的选项说明如下。 • **incoming**：二选一选项，指定携带入接口 IP 地址。 • **incoming-with-label**：二选一选项，指定携带入接口 IP 地址和给上游分配的标签。 • **routerid**：二选一选项，指定携带 LSR ID。 • **routerid-with-label**：二选一选项，指定携带 LSR ID 和给上游分配的标签。

（续表）

步骤	命令	说明
3	**mpls rsvp-te resv-rro transit { { incoming | incoming-with-label } | { routerid | routerid-with-label } | { outgoing | outgoing-with-label } }*** 例如：[Huawei-mpls] **mpls rsvp-te resv-rro transit routerid-with-label incoming outgoing-with-label**	• **outgoing**：二选一选项，指定携带出接口 IP 地址。 • **outgoing-with-label**：二选一选项，指定携带出接口 IP 地址和给下游分配的标签。 以上选项分成了三个部分：入接口 IP 地址（或同时携带标签）、LSR ID（或同时携带标签）、出接口 IP 地址（或同时携带标签），这三个部分之间可同时多选配置，但每一部分内部都只能二选一。修改以上配置时，只能对新创建的 LSP 生效。缺省情况下，Transit LSP 的 Resv 消息携带的 RRO 对象格式是：入接口 IP 地址和给上游分配的标签（**incoming-with-label**）、LSR ID（**routerid**）和给上游分配的 MPLS 标签以及出接口 IP 地址（**outgoing-with-label**），可用 **undo mpls rsvp-te resv-rro transit** 命令恢复缺省配置
	mpls rsvp-te resv-rro egress { { incoming | incoming-with-label } | { routerid | routerid-with-label } }* 例如：[Huawei-mpls] **mpls rsvp-te resv-rro egress routerid-with-label incoming**	（二选一）在 Egress 节点上配置 Resv 消息携带的 RRO 格式。命令中的选项说明参见 **mpls rsvp-te resv-rro transit** 命令。缺省情况下，Egress LSP RRO 携带 RRO 是：入接口地址和给上游分配的标签（**incoming-with-label**）、LSR ID 和给上游分配的标签（routerid-with-label），可用 **undo mpls rsvp-te resv-rro egress** 命令恢复缺省配置

6.1.7 配置 RSVP 认证

RSVP 是一种 MPLS TE 的信令协议，要在相邻节点间传输 RSVP 消息，也必须先在节点间建立 RSVP 邻居关系。第 5 章介绍到，RSVP 是一种工作在网络层的协议（对应 IP 协议号为 46），但它不用来控制数据报文的传输，也不会依赖 TCP 或 UDP 这类同层次的传输协议来进行 RSVP 邻居关系的建立，而是使用 RawIP 协议进行邻居关系的建立。而 RawIP 本身不提供安全性，很容易遭受欺骗攻击，非法路由器在没有授权的情况下和本路由器建立邻居关系，或通过伪造 RSVP 报文的方式和本路由器建立 RSVP 邻居后对路由器进行攻击（比如恶意预留大量的带宽），RSVP 认证可以用来解决此问题。

说明 RawIP 对应的是 Socket 中的 SOCK_RAW（原生套接字），是一种原生数据协议，是建立在 IP 协议之上的传输层原生协议。RawIP 协议可以直接访问 IP 协议，但它是不可靠的，即没有任何控制能确定 RawIP 数据报是否已被接收。

RSVP 认证通过密钥验证的方式来防止欺骗攻击，收到过时报文后认证关系终止。但是这种密钥验证不能防止回放攻击（也称“重放攻击”），就是当非法路由器反复给 RSVP 邻居发送过时的报文（序列号小于当前保存的序列号），会造成 RSVP 认证关系终止，原来已经建立的 CR-LSP 会被拆除。即 RSVP 认证无法解决因 RSVP 报文的失序导致邻居之间认证关系终止的问题，RSVP 认证增强功能就由此诞生。

1. RSVP 认证增强功能实现原理

RSVP 认证增强是在基于原有 RSVP 认证的密钥验证基础上增加了认证生存时间、认证握手机制和消息滑窗特性，使得 RSVP 自身的安全性得到很大的提高，同时加强了在网络阻塞等恶劣网络环境时对邻居关系的合法性进行验证的能力。故总体来说，RSVP

认证增强功能包括以下四方面的功能。

（1）密钥验证

RSVP 认证的密钥验证功能与其他密钥验证功能原理是一样的，就是在建立邻居时对发送的 Hello 报文携带的密钥与本地配置的密钥进行比对，一致表示通过验证，建立邻居关系，不通过则拒绝建立邻居关系，避免受到欺骗攻击。此时需要在建立邻居关系的两个节点上配置相同的验证密钥。在发送 Hello 报文时，节点使用密钥为报文计算得到一个摘要（通过 HMAC-MD5 算法），摘要信息作为报文的一个对象（Integrity 对象），随着报文一起发送到对端节点。对端节点使用相同的密钥、相同的算法重新计算报文摘要，然后比较两个摘要是否相同，如果相同则接收此 Hello 报文，如果不同则丢弃此报文。

（2）认证生存时间

认证生存时间用来指定 RSVP 邻居关系能够持续存在的时间，主要有以下功能。

■ 当 RSVP 邻居之间不存在 CR-LSP 时可以保持 RSVP 邻居关系，直到 RSVP 认证生存时间超时。但 RSVP 认证时间不影响已存在的 CR-LSP 的状态。

■ 可以防止 RSVP 认证无法终止。例如，设备 RTA 和 RTB 建立好了 RSVP 邻居关系后，如果 RTA 后续发给 RTB 的 RSVP Hello 报文因为受篡改而导致密钥被破坏，RTB 收到报文后发现密钥不正确会将报文丢弃，这样就导致 RTA 不断地给 RTB 发送被损坏的 RSVP Hello 消息而该消息不断地被 RTB 丢弃。但邻居之间的认证关系无法拆除。这种情况下，需要配置认证生存时间，如果邻居之间在生存时间内收到合法的 RSVP Hello 报文，则重置 RSVP 认证生存时间，否则认证时间超时后删除 RSVP 邻居的认证关系。

（3）认证握手机制

在两个 RSVP 邻居之间认证成功后，双方互发握手机制的报文。如果握手成功，双方会把对方发过来的握手报文记录在本端，作为一种状态，标志双方已经握手成功。当本端收到过时报文时，就有如下几种处理方式。

■ 如果过时报文表明发送方未使能握手机制，则直接丢弃该报文。

■ 如果过时报文表明发送方也使能了握手机制，且在本端保存与握手成功过的状态，则也直接丢弃该报文；如果本端没有保存与握手成功的状态，则说明是第一次收到发送方的报文，需要与其进行握手。

（4）消息滑窗

消息滑窗用来保存 RSVP 邻居发送的 RSVP 报文的序列号。当滑窗大小为 1 时保存邻居 RSVP 报文的最大序列号，其他情况则可保存多个邻居 RSVP 报文的序列号。比如当滑窗大小为 10，而邻居 RSVP 报文的最大序列号为 80 时，如果没有发生报文乱序，则滑窗保存的内容为[71，80]之间共 10 个序列号。如果已经发生报文乱序，则将报文重新排序后，记录其中 10 个由大到小依次排序的值。

当消息滑窗不为 1 时，当接收到 RSVP 邻居发送的 RSVP 报文之后，如果发现该报文序列号比消息滑窗中保存的最大序列号还大，或者是以前的一个序列号但是比消息滑窗中最小序列号大且没有保存在消息滑窗中时，该报文即为合法报文。接收该报文后，该报文的序列号将被添加到消息滑窗中，然后处理该报文，如果该报文序列号大于消息滑窗中的最大序列号，还将同时删除消息滑窗中最小的那个序列号。序列号比消息滑窗

中最小序列号还小或者已经存在于消息滑窗中的报文将被丢弃。

2. RSVP 密钥管理方式

RSVP 密钥管理包括以下两种方式。

■ MD5 密钥

用户可以使用明文或者密文的方式输入密钥，密钥算法为 MD5。这种密钥管理方式的特点是。

- 每个协议特性都需要配置自己的密钥，密钥不能共享。
- 每个接口、邻居只能配置一个密钥，要更换密钥必须重新配置。

■ Keychain 密钥

Keychain 是一种增强型加密算法，允许用户定义一组密码，形成一个密码串，并且分别为每个密码指定加解密算法及密码使用的有效时间。在收发报文时，系统会按照用户的配置动态选出一个当前有效的密码，并按照与此密码相匹配的加密解密算法，进行发送时加密和接收时解密报文。此外，系统可以依据密码使用的有效时间，自动完成有效密码的切换，避免了长时间不更改密码而导致密码易破解的问题。

这种密钥管理方式的主要特点包括。

- Keychain 的密码、所使用的加解密算法以及密码使用的有效时间可以单独配置，形成一个 Keychain 配置节点，每个 Keychain 配置节点至少需要配置一个密码，并指定加解密算法。
- Keychain 可以被各个协议特性引用，实现密钥集中管理、多特性共享。

RSVP 支持在接口、邻居下引用 Keychain，并仅支持 HMAC-MD5 算法。

3. RSVP 的认证级别

RSVP 的认证级别分为两种。

■ 面向邻居的认证

该级别的认证是指用户可以根据不同的邻居地址配置认证密钥等信息，RSVP 会针对每个邻居进行单独的认证。它有两种配置方式。

- 以邻居设备的某接口的 IP 地址作为邻居地址进行配置。
- 以邻居设备的 LSR ID 作为邻居地址进行配置。

■ 面向接口的认证

用户在接口上配置认证，RSVP 会根据消息的入接口进行认证处理。

面向邻居的认证优先级高于面向接口的认证。只有在高优先级的认证没有使能的情况下才会进行低优先级的认证处理，一旦高优先级认证没有通过，则丢弃该报文。具体的 RSVP 认证功能配置步骤见表 6-7，可在 MPLS TE 隧道各节点上进行。

表 6-7 **RSVP 认证的配置步骤**

步骤	命令	说明
1	**system-view** 例如：<Huawei> **system-view**	进入系统视图
2	**interface** *interface-type interface-number* 例如：[Huawei] **interface** gigabitethernet 1/0/0	（二选一）进入 MPLS TE 链路的接口视图，在接口视图下配置 RSVP 密钥验证功能，**仅在当前接口下生效且优先级最低**

（续表）

步骤	命令	说明
2	**mpls rsvp-te peer** *ip-address* 例如：[Huawei] **mpls rsvp-te peer** 12.0.0.1	（二选一）进入 MPLS RSVP-TE 邻居视图，在 MPLS RSVP-TE 邻居视图下配置 RSVP 密钥验证功能。参数 *ip-address* 邻居的接口地址，当与其 LSR-ID 不相同时，则该密钥认证是基于邻居接口地址的配置。这种配置方式只对该邻居的该接口生效，具有较高安全性，优先级最高。当 *ip-address* 与邻居 LSR-ID 相同时，则该密钥认证是基于邻居 LSR-ID 的配置，使密钥验证功能在该邻居所有接口上生效，其优先级低于基于邻居接口地址配置的密钥验证的优先级，但仍高于基于接口下的 RSVP 验证配置优先级。 【注意】当采用对端设备的 LSR-ID 作为邻居地址时，需要配置 RSVP 认证的设备上必须使能 CSPF 功能。 缺省情况下，没有建立 RSVP 邻居节点，可用 **undo mpls rsvp-te peer** *ip-address* 命令删除 RSVP 邻居节点
3	**mpls rsvp-te authentication** { { **cipher** | **plain** } *auth-key* | **keychain** *keychain-name* } 例如：[Huawei-GigabitEthernet1/0/0] **mpls rsvp-te authentication keychain** kc1 或 [Huawei-mpls-rsvp-te-peer-12.0.0.1] **mpls rsvp-te authentication keychain** kc1	在接口视图下或者 MPLS RSVP-TE 邻居视图下配置 RSVP 验证密钥。命令中的参数和选项说明如下。 • **cipher**：二选一选项，配置 HMAC-MD5 认证，并使用密文方式显示认证密钥。 • **plain**：二选一选项，配置 HMAC-MD5 认证，并使用明文方式显示认证密钥。 • *auth-key*：指定密码字符串，字符串形式，区分大小写，不支持空格。以明文方式输入时，长度范围是 1～255；以 MD5 密文方式输入时，长度范围是 20～392。当输入的字符串两端使用双引号时，可在字符串中输入空格。 • **keychain**：二选一选项，配置 Keychain 认证，引用全局配置的 Keychain，目前只支持 HMAC-MD5 方式。 • *keychain-name*：指定引用的 Keychain 的名称，通过 **keychain** *keychain-name* **mode** { **absolute** | **periodic** { **daily** | **weekly** | **monthly** | **yearly** } }命令定义。 缺省情况下，RSVP 认证功能处于未使能状态，可用 **undo mpls rsvp-te authentication** 命令在接口视图下或者 MPLS RSVP-TE 邻居视图下去使能 RSVP 认证功能
4	**mpls rsvp-te authentication lifetime** *lifetime* 例如：[Huawei-GigabitEthernet1/0/0] **mpls rsvp-te authentication lifetime** 00:20:00 或 [Huawei-mpls-rsvp-te-peer-12.0.0.1] **mpls rsvp-te authentication lifetime** 00:40:00	（可选）设置 RSVP 认证生存时间。参数 *lifetime* 表示认证生存时间，HH：MM:SS 格式，取值范围是 00:00:01～23:59:59，缺省值为 00:30:00，即 30min。 RSVP 认证时间的功能是：当 RSVP 邻居之间不存在 CR-LSP 时可以保持 RSVP 邻居关系，直到 RSVP 认证生存时间超时。配置 RSVP 认证生存时间还可防止 RSVP 认证无法终止的现象。缺省情况下，RSVP-TE 认证生存时间为 30min，可用 **undo mpls rsvp-te authentication lifetime** 命令恢复缺省配置

（续表）

步骤	命令	说明
5	**mpls rsvp-te authentication handshake** 例如：[Huawei-GigabitEthernet1/0/0] **mpls rsvp-te authentication handshake** 或 例如：[Huawei-mpls-rsvp-te-peer-12.0.0.1] **mpls rsvp-te authentication handshake**	（可选）配置 RSVP-TE 握手机制，防止回放攻击，增强网络的安全性。 在本端配置握手机制后，如果本端收到一个没有与自己建立 RSVP 认证关系的邻居所发送的 RSVP 消息，则本端会发送携带本端标识信息的 Challenge 消息给该邻居，邻居收到 Challenge 消息后会向本端回应一个 Response 消息。Response 消息中携带了收到的 Challenge 消息的标识信息，如果其标识信息与本端一致，就确定可以与该邻居建立 RSVP 认证关系。 缺省情况下，没有配置 RSVP-TE 握手机制，可用 **undo mpls rsvp-te authentication handshake** 命令删除 RSVP-TE 握手机制的配置。 **【注意】**如果在邻居之间配置了握手功能，并且需要执行 **mpls rsvp-te authentication lifetime** *lifetime* 命令配置认证生存时间，那么认证生存时间的时长需要大于在 6.1.3 节 **mpls rsvp-te timer refresh** *refresh-interval* 命令配置的 RSVP 刷新消息的发送时间间隔，否则可能造成在认证生存时间内收不到 RSVP 刷新消息而删除认证关系。这样，当下一个刷新消息到来的时候就需要重新进行握手机制的检测，如此反复可能会造成 TE 隧道无法建立或者被删除
6	**mpls rsvp-te authentication window-size** *window-size* 例如：[Huawei-GigabitEthernet1/0/0] **mpls rsvp-te authentication window-size** 10 或 [Huawei-mpls-rsvp-te-peer-12.0.0.1] **mpls rsvp-te authentication window-size** 10	（可选）配置 message window 功能，即指定本地设备可保存的邻居 RSVP 消息有效序列号的个数，避免因 RSVP 报文的失序导致邻居之间认证关系终止。参数 *window-size* 用来指定消息滑窗大小，整数形式，取值范围是 1～64。当配置的 window-size 大于 1 时，本地就可以保存邻居 RSVP 消息的最近多个有效序列号。 缺省情况下，RSVP 消息窗口大小是 1，即本地设备只能保存邻居 RSVP 消息的一个最近的最大的序列号，可用 **undo mpls rsvp-te authentication window-size** 命令恢复缺省配置。 **【注意】**当 RSVP 接口类型为 Eth-Trunk 时，RSVP 邻居之间只在 Trunk 链路上建立一个邻居关系。RSVP 消息可以从 Trunk 的任意一个成员接口接收，且各个成员口不是按顺序接收报文的，这样可能造成 RSVP 消息失序，因此必须配置 RSVP 消息滑动窗口。建议将滑动窗口大小配置为大于 32。如果滑动窗口设置过小，收到失序的 RSVP 消息有些可能不在窗口范围内而被丢弃，这样会导致 RSVP 邻居关系终止
7	**quit** 例如：[Huawei-GigabitEthernet1/0/0] **quit** 或 [Huawei-mpls-rsvp-te-peer-12.0.0.1] **quit**	返回系统视图

（续表）

步骤	命令	说明
8	**mpls** 例如：[Huawei] **mpls**	进入 MPLS 视图
9	**mpls rsvp-te retrans-timer challenge** *retransmission-interval* 例如：[Huawei-mpls] **mpls rsvp-te retrans-timer challenge** 800	配置 challenge 消息重传间隔整数形式，取值范围是 500～10000，单位是毫秒。 当两个节点间的认证消息失序以后，一个节点将向另一个节点发送 challenge 消息请求恢复连接；如果没有收到对端的响应消息，该节点将在一个重传间隔后重传 challenge 消息。 缺省情况下，challenge 消息的重传间隔为 1000ms，可用 **undo mpls rsvp-te retrans-timer challenge** 命令恢复缺省配置
10	**mpls rsvp-te challenge-lost** *max-miss-times* 例如：[Huawei-mpls] **mpls rsvp-te challenge-lost** 5	配置 RSVP 认证过程中认证方允许重传 challenge 消息的最大次数，整数形式，取值范围是 1～10。 如果达到最大重传次数后，仍未收到对方的响应消息，则认证失败；如果达到最大次数前，认证成功，则清零 challenge 消息重传计数。 缺省情况下，challenge 消息的最大重传次数为 3，可用 **undo mpls rsvp-te challenge-lost** 命令恢复缺省值

6.1.8 RSVP-TE 参数调整管理

已经完成上述 RSVP 信令参数调整的配置后，可通过以下 display 命令进行配置检查或结果验证。

- **display mpls rsvp-te**：查看 RSVP-TE 的相关信息。
- **display default-parameter mpls rsvp-te**：查看 MPLS RSVP-TE 缺省参数信息。
- **display mpls rsvp-te session** *ingress-lsr-id tunnel-id egress-lsr-id*：查看指定 RSVP 会话的所有信息。
- **display mpls rsvp-te psb-content** [*ingress-lsr-id tunnel-id lsp-id*]：查看 RSVP-TE PSB 信息。
- **display mpls rsvp-te rsb-content** [*ingress-lsr-id tunnel-id lsp-id*]：查看 RSVP-TE RSB 信息。
- **display mpls rsvp-te statistics** { **global** | **interface** [*interface-type interface-number*] }：查看 RSVP-TE 运行统计信息。
- **display mpls rsvp-te peer** [**interface** *interface-type interface-number*]：查看使能 RSVP-TE 的接口的 RSVP-TE 邻居信息。

6.1.9 RSVP 认证配置示例

如图 6-2 所示，LSRA 和 LSRB 之间的 Eth-Trunk 1 的成员接口为 GE1/0/0 和 GE2/0/0。使用 RSVP 建立了一条从 LSRA 到 LSRC 的 MPLS TE 隧道。现要求配置 Handshake（握手）功能使LSRA和LSRB之间互相进行RSVP密钥认证，防止回放攻击以及伪造的RSVP资源预留请求非法占用网络资源，并配置 message window（消息窗口）功能解决 RSVP

报文失序问题。

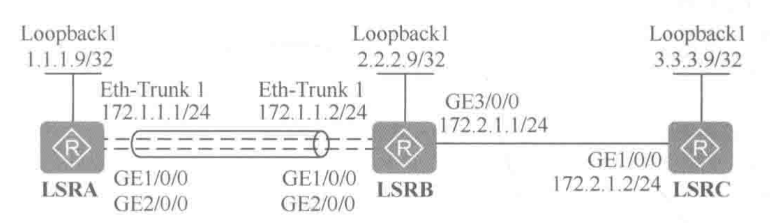


图 6-2　RSVP 认证配置示例的拓扑结构

1. 基本配置思路分析

虽然本示例介绍的是 RSVP 认证配置，但这项功能是在建立 MPLS TE 或 MPLS DS-TE 隧道基础上配置的，所以在配置 RSVP 认证功能之前仍要先完成 LSRA 与 LSRC 之间 CR-LSP 建立的配置。

结合第 5 章介绍的基本 MPLS TE 隧道功能的配置方法，以及 6.1.7 节介绍的 RSVP 认证功能的配置方法可得出本示例的基本配置思路如下。

（1）配置各节点接口（包括 Loopback 接口）的 IP 地址，并使用 OSPF 协议实现各节点之间公网路由可达。

（2）配置各节点的 LSR ID，并全局使能各节点以及公网接口的 MPLS、MPLS TE、MPLS RSVP-TE 功能，在 LSRA 上使能 CSPF。

（3）在入节点创建隧道接口，指定隧道的 IP 地址、隧道协议、目的地址、隧道 ID、动态信令协议 RSVP-TE。本示例仅介绍以 LSRA 为入节点，到达 LSRC 的 MPLS TE 隧道的配置方法。

（4）在各节点上配置 OSPF TE，使能 OSPF 的 Opaque 能力，以使 OSPF 支持 MPLS 信息发布。

（5）在 LSRA 和 LSRB 上配置 RSVP 消息密钥验证功能，并使能 Handshake 功能，防止伪造的 RSVP 资源预留请求非法占用网络资源，配置 message window 为 32，解决可能发生的 RSVP 报文失序问题。

2. 具体配置步骤

（1）配置各接口的 IP 地址，并配置 OSPF 协议，实现 MPLS TE 公网三层互通。

LSRA 上的配置。

因为 LSRA 与 LSRB 之间是通过 Eth-Trunk 链路连接的，所以要先配置 Eth-Trunk 聚合链路，把聚合链路接口转换成三层模式，配置 IP 地址和 OSPF 协议。

```
<Huawei> system-view
[Huawei] sysname LSRA
[LSRA] interface eth-trunk 1     #---创建 Eth-Trunk 聚合链路
[LSRA-Eth-Trunk1] undo portswitch   #---转换成三层模式
[LSRA-Eth-Trunk1] ip address 172.1.1.1 255.255.255.0
[LSRA-Eth-Trunk1] quit
[LSRA] interface gigabitethernet 1/0/0
[LSRA-GigabitEthernet1/0/0] eth-trunk 1     #---把 GE1/0/0 接口加入聚合链路
[LSRA-GigabitEthernet1/0/0] quit
[LSRA] interface gigabitethernet 2/0/0
[LSRA-GigabitEthernet2/0/0] eth-trunk 1
[LSRA-GigabitEthernet2/0/0] quit
```

```
[LSRA] interface loopback 1
[LSRA-LoopBack1] ip address 1.1.1.9 255.255.255.255
[LSRA-LoopBack1] quit
[LSRA] ospf 1
[LSRA-ospf-1] area 0
[LSRA-ospf-1-area-0.0.0.0] network 1.1.1.9 0.0.0.0
[LSRA-ospf-1-area-0.0.0.0] network 172.1.1.0 0.0.0.255
[LSRA-ospf-1-area-0.0.0.0] quit
[LSRA-ospf-1] quit
```

LSRB 的配置。

因为 LSRB 与 LSRA 之间是通过 Eth-Trunk 链路连接的，所以也要先配置 Eth-Trunk 聚合链路，把聚合链路接口转换成三层模式，配置 IP 地址和 OSPF 协议。

```
<Huawei> system-view
[Huawei] sysname LSRB
[LSRB] interface eth-trunk 1     #---创建 Eth-Trunk 聚合链路
[LSRB-Eth-Trunk1] undo portswitch    #---转换成三层模式
[LSRB-Eth-Trunk1] ip address 172.1.1.2 255.255.255.0
[LSRB-Eth-Trunk1] quit
[LSRB] interface gigabitethernet 1/0/0
[LSRB-GigabitEthernet1/0/0] eth-trunk 1     #---把 GE1/0/0 接口加入聚合链路
[LSRB-GigabitEthernet1/0/0] quit
[LSRB] interface gigabitethernet 2/0/0
[LSRB-GigabitEthernet2/0/0] eth-trunk 1
[LSRB-GigabitEthernet2/0/0] quit
[LSRB] interface gigabitethernet 3/0/0
[LSRB- GigabitEthernet3/0/0] ip address 172.2.1.1 255.255.255.255
[LSRB- GigabitEthernet3/0/0] quit
[LSRB] interface loopback 1
[LSRB-LoopBack1] ip address 2.2.2.9 255.255.255.255
[LSRB-LoopBack1] quit
[LSRB] ospf 1
[LSRB-ospf-1] area 0
[LSRB-ospf-1-area-0.0.0.0] network 2.2.2.9 0.0.0.0
[LSRB-ospf-1-area-0.0.0.0] network 172.1.1.0 0.0.0.255
[LSRB-ospf-1-area-0.0.0.0] network 172.2.1.0 0.0.0.255
[LSRB-ospf-1-area-0.0.0.0] quit
[LSRB-ospf-1] quit
```

LSRC 的配置。

```
<Huawei> system-view
[Huawei] sysname LSRC
[LSRC] interface gigabitethernet 1/0/0
[LSRC- GigabitEthernet1/0/0] ip address 172.2.1.2 255.255.255.255
[LSRC- GigabitEthernet1/0/0] quit
[LSRC] interface loopback 1
[LSRC-LoopBack1] ip address 3.3.3.9 255.255.255.255
[LSRC-LoopBack1] quit
[LSRC] ospf 1
[LSRC-ospf-1] area 0
[LSRC-ospf-1-area-0.0.0.0] network 3.3.3.9 0.0.0.0
[LSRC-ospf-1-area-0.0.0.0] network 172.2.1.0 0.0.0.255
[LSRC-ospf-1-area-0.0.0.0] quit
[LSRC-ospf-1] quit
```

以上配置完成后，在各节点上执行 **display ip routing-table** 命令，可以看到相互之间

都学习到了到对方 Loopback1 的路由。

（2）在各节点上配置 MPLS LSR ID，并在全局和各公网接口上使能 MPLS、MPLS TE 和 MPLS RSVP-TE 能力，在隧道入节点 LSRA 上全局使能 CSPF 能力。

LSRA 上的配置。

```
[LSRA] mpls lsr-id 1.1.1.9
[LSRA] mpls
[LSRA-mpls] mpls te
[LSRA-mpls] mpls rsvp-te
[LSRA-mpls] mpls te cspf
[LSRA-mpls] quit
[LSRA] interface eth-trunk 1
[LSRA-Eth-Trunk1] mpls
[LSRA-Eth-Trunk1] mpls te
[LSRA-Eth-Trunk1] mpls rsvp-te
[LSRA-Eth-Trunk1] quit
```

LSRB 上的配置。

```
[LSRB] mpls lsr-id 2.2.2.9
[LSRB] mpls
[LSRB-mpls] mpls te
[LSRB-mpls] mpls rsvp-te
[LSRB-mpls] quit
[LSRB] interface eth-trunk 1
[LSRB-Eth-Trunk1] mpls
[LSRB-Eth-Trunk1] mpls te
[LSRB-Eth-Trunk1] mpls rsvp-te
[LSRB-Eth-Trunk1] quit
[LSRB] interface gigabitethernet 3/0/0
[LSRB-GigabitEthernet3/0/0] mpls
[LSRB-GigabitEthernet3/0/0] mpls te
[LSRB-GigabitEthernet3/0/0] mpls rsvp-te
[LSRB-GigabitEthernet3/0/0] quit
```

LSRC 上的配置。

```
[LSRC] mpls lsr-id 3.3.3.9
[LSRC] mpls
[LSRC-mpls] mpls te
[LSRC-mpls] mpls rsvp-te
[LSRC-mpls] quit
[LSRC] interface gigabitethernet 1/0/0
[LSRC-GigabitEthernet1/0/0] mpls
[LSRC-GigabitEthernet1/0/0] mpls te
[LSRC-GigabitEthernet1/0/0] mpls rsvp-te
[LSRC-GigabitEthernet1/0/0] quit
```

（3）配置 OSPF TE，发布 TE 信息。

LSRA 上的配置。

```
[LSRA] ospf
[LSRA-ospf-1] opaque-capability enable
[LSRA-ospf-1] area 0
[LSRA-ospf-1-area-0.0.0.0] mpls-te enable
[LSRA-ospf-1-area-0.0.0.0] quit
[LSRA-ospf-1] quit
```

LSRB 上的配置。

```
[LSRB] ospf
[LSRB-ospf-1] opaque-capability enable
[LSRB-ospf-1] area 0
[LSRB-ospf-1-area-0.0.0.0] mpls-te enable
[LSRB-ospf-1-area-0.0.0.0] quit
[LSRB-ospf-1] quit
```

LSRC 上的配置。

```
[LSRC] ospf
[LSRC-ospf-1] opaque-capability enable
[LSRC-ospf-1] area 0
[LSRC-ospf-1-area-0.0.0.0] mpls-te enable
[LSRC-ospf-1-area-0.0.0.0] quit
[LSRC-ospf-1] quit
```

（4）在入节点创建并配置 MPLS TE Tunnel 接口。

```
[LSRA] interface tunnel 0/0/1
[LSRA-Tunnel0/0/1] ip address unnumbered interface loopback 1
[LSRA-Tunnel0/0/1] tunnel-protocol mpls te
[LSRA-Tunnel0/0/1] destination 3.3.3.9
[LSRA-Tunnel0/0/1] mpls te tunnel-id 101
[LSRA-Tunnel0/0/1] mpls te commit
[LSRA-Tunnel0/0/1] quit
```

以上配置完成后，在 LSRA 上执行 **display interface tunnel** 命令可以看到隧道接口状态为 Up。

```
[LSRA] display interface tunnel 0/0/1
Tunnel0/0/1 current state : Up
Line protocol current state : Up
Last line protocol Up time : 2013-02-22 14:28:37
Description:...
```

（5）在 LSRA、LSRB 的 MPLS TE 链路的接口上配置 RSVP 验证。假设验证密钥为 Huawei@1234，使能握手功能，并设置消息窗口中大小为 32。

LSRA 上的配置。

```
[LSRA] interface eth-trunk 1
[LSRA-Eth-Trunk1] mpls rsvp-te authentication cipher Huawei@1234 #---配置 RSVP-TE 邻居认证密钥为 Huawei@1234
[LSRA-Eth-Trunk1] mpls rsvp-te authentication handshake #--使能 RSVP-TE 握手功能
[LSRA-Eth-Trunk1] mpls rsvp-te authentication window-size 32 #---配置 RSVP-TE 认证消息滑窗大小为 32
[LSRA-Eth-Trunk1] quit
```

LSRB 上的配置。

```
[LSRB] interface eth-trunk 1
[LSRB-Eth-Trunk1] mpls rsvp-te authentication cipher Huawei@1234
[LSRB-Eth-Trunk1] mpls rsvp-te authentication handshake
[LSRB-Eth-Trunk1] mpls rsvp-te authentication window-size 32
[LSRB-Eth-Trunk1] quit
```

3. 配置结果验证

以上配置完成后，在 MPLS TE 隧道建立成功、LSRA 和 LSRB 之间的 RSVP 邻居关系建立成功后，在 LSRA 或 LSRB 上执行 **display mpls rsvp-te interface** 命令，可查看到有关 RSVP 认证的配置信息。其中显示的"**Num of Neighbors: 1**"就代表 RSVP 邻居建立成功了。

```
[LSRA] display mpls rsvp-te interface eth-trunk 1
Interface: Eth-Trunk1
```

```
Interface Address: 172.1.1.1
Interface state: Up                    Interface Index: 0x36
Total-BW: 0                            Used-BW: 0
Hello configured: NO                   Num of Neighbors: 1
SRefresh feature: DISABLE              SRefresh Interval: 30 sec
Mpls Mtu: 1500                         Retransmit Interval: 5000 msec
Increment Value: 1                     Authentication: ENABLE
Challenge: ENABLE                      WindowSize: 32
Next Seq # to be sent:2767789282 0     Key ID: 0xa4ff1cdc0000
Bfd Enabled: DISABLE                   Bfd Min-Tx: 1000
Bfd Min-Rx: 1000                       Bfd Detect-Multi: 3
```

6.2　调整 CR-LSP 的路径选择

CSPF 使用 TEDB 和约束条件计算出符合要求的路径，并通过信令协议建立 CR-LSP。MPLS TE 提供多种方式影响 CSPF 的计算，从而调整 CR-LSP 的建立，包括：CSPF 的仲裁方法、选路使用的度量、CR-LSP 的跳数限制值、路由锁定、管理组和亲和属性、失效链路定时器等方式。在通过这些配置任务调整 CR-LSP 的路径选择之前，须完成动态 MPLS TE 隧道或动态 DS-TE 隧道配置。

6.2.1　配置 CSPF 的仲裁方法

当入节点在使用 CSPF 计算路径的过程中遇到多条权值相同的路径时，CSPF 也仅会通过仲裁选择其中一条路径，而不会同时选择多条路径。可通过配置 CSPF 仲裁的方法影响 CSPF 的最终路径选择。但要注意的是，CSPF 功能仅在隧道入节点上才需要使能、进行 CR-LSP 路径计算，所以此处的仲裁方法仅需在**入节点**上全局（适用于所有 TE 隧道）或具体的 Tunnel 接口上配置，Tunnel 接口上的配置优先级高于全局配置，具体配置步骤见表 6-8。

表 6-8　　配置 **CSPF** 的仲裁方法的步骤

步骤	命令	说明
1	**system-view** 例如：<Huawei> **system-view**	进入系统视图
2	**mpls** 例如：[Huawei] **mpls**	进入 MPLS 视图
3	**mpls te tie-breaking** { **least-fill** | **most-fill** | **random** } 例如：[Huawei-mpls] **mpls te tie-breaking least-fill**	配置全局的 CSPF 仲裁方法。命令中的选项说明如下。 • **least-fill**：多选一选项，指定优选已使用的可保留带宽占链路最大可预留带宽比值最小的路径。 • **most-fill**：多选一选项，指定优选已使用的可保留带宽占链路最大可预留带宽比值最大的路径。 • **random**：多选一选项，指定随机选择路径。 链路最大可预留带宽是指通过 **mpls te bandwidth max-reservable-bandwidth** *bw-value* 接口视图命令配置的

（续表）

步骤	命令	说明
3	**mpls te tie-breaking** { **least-fill** | **most-fill** | **random** } 例如：[Huawei-mpls] **mpls te tie-breaking least-fill**	【注意】在比值相同的情况下，例如都没有利用保留带宽，或者利用的份额都是一样，此时不管配置的是 **least-fill** 还是 **most-fill** 选项，都会选择首先发现的链路。 缺省情况下，仲裁方法为随机选择方式 **random**，可用 **undo mpls te tie-breaking** 命令恢复缺省设置
4	**quit** 例如：[Huawei-mpls] **quit**	返回系统视图
5	**interface tunnel** *interface-number* 例如：[Huawei] **interface tunnel** 0/0/1	进入 MPLS TE 隧道的 Tunnel 接口视图
6	**mpls te tie-breaking** { **least-fill** | **most-fill** | **random** } 例如：[Huawei-Tunnel0/0/1] **mpls te tie-breaking least-fill**	配置当前 Tunnel 接口的 CSPF 仲裁方法。命令中的选项说明参见本表第 3 步。 隧道优先使用本隧道接口视图下配置的仲裁方法；如果隧道接口视图下没有配置，将使用 MPLS 视图下配置的全局仲裁方法
7	**mpls te commit** 例如：[Huawei-Tunnel0/0/1] **mpls te commit**	提交隧道当前配置。凡发生了 TE 隧道配置更改，均要执行本命令，才能使配置更改生效

6.2.2 配置选路使用的度量

通过配置选路使用的度量，可以在**入节点**上指定 TE 隧道或链路使用的度量类型，影响 CR-LSP 路径的选择。可以在 MPLS 视图下全局配置（适用于所有 TE 隧道）或具体的 Tunnel 接口下配置度量类型，Tunnel 接口下配置优先；如果采用 TE 类型的度量，还可选在具体的物理接口视图下配置 TE 度量值，具体配置步骤见表 6-9。

表 6-9 配置 TE 隧道或链路使用的度量类型的步骤

步骤	命令	说明
1	**system-view** 例如：<Huawei> **system-view**	进入系统视图
2	**mpls** 例如：[Huawei] **mpls**	进入 MPLS 视图
3	**mpls te path metric-type** { **igp** | **te** } 例如：[Huawei-mpls] **mpls te path metric-type igp**	全局配置 CR-LSP 选路时使用的度量类型，命令中的选项说明如下。 • **igp**：二选一选项，指定使用 IGP（如 OSPF 或 IS-IS）度量。 • **te**：二选一选项，指定使用 TE 度量。 【说明】如果在隧道接口视图下没有按本表第 6 步配置度量类型，则使用此处 MPLS 视图下的度量类型；否则使用隧道接口视图下的度量类型。 缺省情况下，TE 隧道选路时使用的度量类型为 TE，可用 **undo mpls te path metric-type** 命令恢复缺省设置
4	**quit** 例如：[Huawei-mpls] **quit**	返回系统视图

（续表）

步骤	命令	说明
5	**interface tunnel** *interface-number* 例如：[Huawei] **interface tunnel** 0/0/1	进入 MPLS TE 隧道的 Tunnel 接口视图
6	**mpls te path metric-type** { **igp** | **te** } 例如：[Huawei-Tunnel0/0/1] **mpls te path metric-type igp**	配置当前隧道选路时使用的度量类型。命令中的选项参见本表第 3 步说明。 如果在 Tunnel 接口下配置了度量类型，最终以具体 Tunnel 接口下配置的为准，否则采用在第 3 步中全局配置。 缺省情况下，隧道选路时使用链路的 TE 度量，可用 **undo mpls te path metric-type** 命令恢复缺省设置
7	**mpls te commit** 例如：[Huawei-Tunnel0/0/1] **mpls te commit**	提交隧道当前配置。凡发生了 TE 隧道配置更改，均要执行本命令，才能使配置更改生效
8	**quit** 例如：[Huawei-Tunnel0/0/1] **quit**	返回系统视图
9	**interface** *interface-type interface-number* 例如：[Huawei] **interface** gigabitethernet 1/0/0	进入 MPLS TE 链路的接口视图
10	**mpls te metric** *value* 例如：[Huawei-GigabitEthernet1/0/0] **mpls te metric** 20	（可选）当 TE 隧道采用 TE 度量类型时，可配置链路的 TE 度量值，整数形式，取值范围是 1～16777215。 缺省情况下，链路使用其 IGP 度量值作为 TE 的度量值，可用 **undo mpls te metric** 命令恢复缺省配置。 **【注意】**当 IGP 协议为 OSPF 且当前设备的状态为 stub router 时，本命令不生效

6.2.3 配置 CR-LSP 的跳数限制值

跳数限制值作为 CR-LSP 建立时的选路条件之一，就像管理组与亲和属性一样，可以限制一条 CR-LSP 允许选择的路径跳数不超过某个值，最终也可影响 CR-LSP 路径的选择。可在入节点上按表 6-10 的步骤在具体的 Tunnel 接口下配置对应 CR-LSP 的跳数限制值。

表 6-10　配置 CR-LSP 的跳数限制值的步骤

步骤	命令	说明
1	**system-view** 例如：<Huawei> **system-view**	进入系统视图
2	**interface tunnel** *interface-number* 例如：[Huawei] **interface tunnel** 0/0/1	进入 MPLS TE 隧道的 Tunnel 接口视图
3	**mpls te hop-limit** *hop-limit-value* [**best-effort** | **secondary**] 例如：[Huawei-Tunnel0/0/1] **mpls te hop-limit** 10 **best-effort**	限制该 Tunnel 的 CR-LSP 的路径跳数。命令中的参数和选项说明如下。 • *hop-limit-value*：指定跳数限制值，整数形式，取值范围是 1～32

（续表）

步骤	命令	说明
3	**mpls te hop-limit** *hop-limit-value* [best-effort | secondary] 例如：[Huawei-Tunnel0/0/1] **mpls te hop-limit** 10 **best-effort**	• **best-effort**：二选一可选项，指定以上参数 *hop-limit-value* 值为逃生路径的最大跳数。逃生路径是指在主、备 CR-LSP 都故障时，创建一条临时的 CR-LSP，将业务流量切换到逃生路径上。有关 CP-LSP 备份的配置将在第 7 章介绍 • **secondary**：二选一可选项，指定以上参数 *hop-limit-value* 值为备份路径的最大跳数。 如果不选择 **best-effort** | **secondary** 可选项，则参数所配置的跳数限制值是针对主 CR-LSP 路径而言的。 缺省情况下，CR-LSP 的路径跳数限制值为 32，可用 **undo mpls te hop-limit** [**best-effort** | **secondary**]命令恢复对应路径的跳数限制值为缺省值
4	**mpls te commit** 例如：[Huawei-Tunnel0/0/1] **mpls te commit**	提交隧道当前配置。凡发生了 TE 隧道配置更改，均要执行本命令，才能使配置更改生效

6.2.4 配置路由锁定

通过在 Tunnel 接口下配置路由锁定，CR-LSP 能够保持初始选定的路径，而不会按照新的可能的路径重新建立，这样可以保持 CR-LSP 路径的稳定性。可在**入节点**上按照表 6-11 的步骤配置路由锁定功能，不能同时使用 CR-LSP 重优化。

表 6-11 配置路由锁定功能的步骤

步骤	命令	说明
1	**system-view** 例如：<Huawei> **system-view**	进入系统视图
2	**interface tunnel** *interface-number* 例如：[Huawei] **interface tunnel** 0/0/1	进入 MPLS TE 隧道的 Tunnel 接口视图
3	**mpls te route-pinning** 例如：[Huawei-Tunnel0/0/1] **mpls te route-pinning**	在以上隧道接口上使能路由锁定功能，使其总是采用初始的路径选择，不进行重新的路径计算。 缺省情况下，路由锁定功能未使能，可用 **undo mpls te route-pinning** 命令去使能路由锁定功能
4	**mpls te commit** 例如：[Huawei-Tunnel0/0/1] **mpls te commit**	提交隧道当前配置。凡发生了 TE 隧道配置更改，均要执行本命令，才能使配置更改生效

6.2.5 配置管理组与亲和属性

进行管理组属性配置，可以影响新创建的 CR-LSP 的路径选择；而进行亲和属性配置，可以影响该隧道已建立的 CR-LSP 的路径选择，系统将为该隧道重新计算路径。可按照表 6-12 的步骤配置管理组与亲和属性。

表 6-12 配置管理组和亲和属性的步骤

步骤	命令	说明
1	**system-view** 例如：<Huawei> **system-view**	进入系统视图
2	**interface** *interface-type interface-number* 例如：[Huawei] **interface** gigabitethernet 1/0/0	进入 MPLS TE 链路的接口视图，**需要在存在多条路径的各节点的出接口配置下的链路管理组属性**
3	**mpls te link administrative group** *value* 例如：[Huawei-GigabitEthernet1/0/0] **mpls te link administrative group** 101	配置链路的管理组。参数 *value* 用来指定链路属性，在选路时与隧道的 affinity（亲和属性）位进行比较，十六进制形式，取值范围是 0x0～0xFFFFFFFF，表示 32 个属性，每个属性占一位。 **【说明】**这个链路管理组的链路属性配置不是随意的，必须与对应的 Tunnel 接口下的亲和属性配置同步规划，因为如果要使某条链路被对应 TE 隧道选用，在所有掩码为 1 的位中，管理组中至少有 1 位与亲和属性中的相应位都为 1；且亲和属性为 0 的位，对应的管理组属性位不能为 1。 接口属性将在全局范围内扩散，从而可以用作隧道源端的路径选择标准。**更改链路的管理组属性配置，仅对新创建的 CR-LSP 生效，不影响已建立的 CR-LSP。** 缺省情况下，链路管理属性的值为 0x0，可用 **undo mpls te link administrative group** 命令恢复缺省值
4	**quit** 例如：[Huawei-GigabitEthernet1/0/0] **quit**	返回系统视图
5	**interface tunnel** *interface-number* 例如：[Huawei] **interface tunnel** 0/0/1	进入 MPLS TE 隧道的 Tunnel 接口视图。**仅需在入节点上配置 Tunnel 接口亲和属性**
6	**mpls te affinity property** *properties* [**mask** *mask-value*] [**secondary** | **best-effort**] 例如：[Huawei-Tunnel0/0/1] **mpls te affinity property** a04 **mask** e0c	配置隧道的亲和属性。**更改隧道的亲和属性配置会影响该隧道已建立的 CR-LSP，系统将为该隧道重新计算路径**。命令中的参数和选项说明如下。 • *Properties*：指定 MPLS TE 隧道使用的链路亲和属性，十六进制形式，取值范围是 0x0～0xFFFFFFFF，每一位代表一种属性。 • **mask** *mask-value*：可选参数，掩码，指定需要检查的链路管理组属性位，十六进制形式，取值范围是 0x0～0xFFFFFFFF，每一位代表一种属性。掩码决定了设备需要检查哪些链路管理组属性。 • **secondary**：二选一可选项，指定以上配置的亲和属性和掩码是针对备份 CR-LSP。 • **best-effort**：二选一可选项，指定以上配置的亲和属性和掩码是针对逃生路径。 如果不选择 **secondary** | **best-effort** 可选项，则配置的亲和属性的掩码是针对主 CR-LSP，有关备份 CR-LSP 和逃生路径的配置将在第 7 章介绍。 缺省情况下，链路管理组的值为 0x0；隧道的亲和属性为 0x0，掩码 0x0，可用 **undo mpls te affinity property** *properties* [**mask** *mask-value*] { **secondary** | **best-effort** } 命令恢复缺省值

（续表）

步骤	命令	说明
7	**mpls te commit** 例如：[Huawei-Tunnel0/0/1] **mpls te commit**	提交隧道当前配置。凡发生了 TE 隧道配置更改，均要执行本命令，才能使配置更改生效

6.2.6 配置 CR-LSP 和 Overload 联动功能

Overload 是 IS-IS LSP（链路状态 PDU）中的一个标志位，则当其他 IS-IS 邻居设备收到 Overload 标志位置 1 的 LSP 时，就知道发送该 LSP 的设备处于超载状态，这样在计算 IS-IS 路由时会绕开该设备。

在 MPLS TE 或 MPLS DS-TE 中可使用 IS-IS TE 进行 TE 消息发布，然后通过 CSPF 进行 CR-LSP 路径计算，所以 IS-IS 中 LSP 的 Overload 标志位可影响 CR-LSP 路径的选择。

有两种方式可以使某节点成为 Overload 节点。

■ 当某节点在网络中由于承载业务较多，出现了超负荷工作状态，导致系统资源耗尽时，该节点会标志自己为 Overload 节点。这是一种自动方式。

■ 当网络管理员发现网络中某节点（**只能是中间节点**）承载业务较多，出现 CPU 比较繁忙的状态时，可以通过执行 **set-overload** 命令，标志该节点为 Overload 节点。

在部署 MPLS TE 业务时，如果希望 TE 流量避开 Overload 节点，即新建立的 CR-LSP 不经过 Overload 节点，可以在**入节点**上按照表 6-13 的步骤配置 CR-LSP 和 Overload 的联动功能。这样可以减轻 Overload 节点的压力，同时也提高 CR-LSP 的可靠性。

表 6-13 配置 CR-LSP 和 Overload 联动功能的步骤

步骤	命令	说明
1	**system-view** 例如：<Huawei> **system-view**	进入系统视图
2	**interface tunnel** *interface-number* 例如：[Huawei] **interface tunnel** 0/0/1	进入 MPLS TE 隧道的 Tunnel 接口视图
3	**mpls te record-route [label]** 例如：[Huawei-Tunnel0/0/1] **mpls te record-route label**	使能建立隧道时记录路由和标签功能。 在没有配置显式路径的情况下，MPLS TE Tunnel 建立成功后，系统不会记录隧道的详细路径。如果需要查看隧道的详细路径，可以在 Tunnel 接口下执行该命令，在 Path、Resv 消息中携带 RRO，记录消息经过每一跳的 IP 地址。如果选择了可选项 **label**，则同时记录经过每一跳的标签。 【说明】如果网络规模比较大，不建议使用此命令。因为 RRO 会记录每一跳的 IP 地址，如果跳数很多，则会导致 Path 或 Resv 消息过大，降低系统性能。 缺省情况下，未使能建立隧道时记录路由和标签功能，可用 **undo mpls te record-route** 命令去使能建立隧道时记录路由和标签功能
4	**quit** 例如：[Huawei-Tunnel0/0/1] **quit**	返回系统视图

（续表）

步骤	命令	说明
5	**mpls** 例如：[Huawei] **mpls**	进入 MPLS 视图
6	**mpls te path-selection overload** 例如：[Huawei-mpls] **mpls te path-selection overload**	配置 CR-LSP 和 IS-IS Overload 联动功能，使能 CSPF 在计算路由的时候排除 IS-IS Overload（过载）节点，从而使流量避开 Overload 节点。执行本命令后。 • 已经建立好的 CR-LSP 会进行重优化，CSPF 会重新算路，使流量避开 Overload 节点。 • 对于新建的 LSP，则 CSPF 在进行路径计算时就会排除网络中的 Overload 节点，使流量避开 Overload 节点。 【注意】对于具体 CR-LSP 来说，如果在 Ingress 节点或者 Egress 节点上通过 IS-IS 进程视图下的 **set-overload** [**on-startup** [*timeout1* | **start-from-nbr** *system-id* [*timeout1* [*timeout2*]] | **wait-for-bgp** [*timeout1*]] [**send-sa-bit** [*timeout3*]]] [**allow** { **interlevel** | **external** }*]命令被设置为 Overload 节点，那么执行本命令对已经建立的 CR-LSP 不生效，即该 CR-LSP 不会进行重优化，而新的 CR-LSP 也无法建立。 缺省情况下，没有配置 CR-LSP 和 IS-IS Overload 联动功能，可用 **undo mpls te path-selection** 命令恢复为缺省配置

6.2.7　配置失效链路定时器

CSPF 是依据本地维护的 TEDB 来计算到目的地址最短路径的，然后信令协议根据 CSPF 计算路径请求和预留资源。但如果网络某链路发生故障，路由协议有时可能没有及时通知 CSPF 更新 TEDB，导致信令协议从 CSPF 得到的路径可能包含存在故障的链路。因为当链路发生故障时，信令协议的控制报文（如 RSVP 的 Path 消息）将会丢失，信令协议将返回错误消息通知上游节点。上游节点收到链路错误消息时，触发 CSPF 重新计算路径。但又因为 TEDB 没有更新，CSPF 重计算并告知信令协议的路径仍然包含了故障的链路，这样一来信令协议的控制报文又将被丢弃，信令协议又返回错误消息，触发 CSPF 重新计算路径。如此反复，直到 TEDB 得到更新。

为了避免上述情况的发生，当收到信令协议返回错误消息通知链路故障时，CSPF 将故障链路的状态置为 INACTIVE（无效），并启动失效链路定时器。这样故障链路将不再参与 CSPF 的路径计算，直到 CSPF 收到数据库更新事件或者链路失效定时器超时。在链路失效定时器超时前，如果收到数据库更新事件，CSPF 将删除链路失效定时器。

配置失效链路定时器的步骤是在**入节点**上按照表 6-14 的步骤进行配置，因为 CR-LSP 路径的计算都是在入节点进行的。

表 6-14　配置失效链路定时器的步骤

步骤	命令	说明
1	**system-view** 例如：<Huawei> **system-view**	进入系统视图

（续表）

步骤	命令	说明
2	**mpls** 例如：[Huawei] **mpls**	进入 MPLS 视图
3	**mpls te cspf timer failed-link** *interval* 例如：[Huawei-mpls] **mpls te cspf timer failed-link** 50	配置 CSPF 失效链路定时器，整数形式，取值范围是 1～300，单位为秒。执行此命令之前，必须先使用 **mpls te cspf** 命令使能 CSPF 功能。 一旦一条链路状态变为 Down，失效链路定时器就会启动。如果 IGP 在定时器超时之前删除或修改此链路，IGP 将会把删除或修改情况通知 CSPF。CSPF 在 TEDB 中更新链路，并停止定时器；如果定时器超时，IGP 还没有删除链路时，链路状态将被更新为 Up。 【注意】失效链路定时器的值只具有本地意义，如果各个节点上配置的定时器值不同，则可能存在同一条链路在某些节点上的状态是 ACTIVE（有效），而另一些节点上是 INACTIVE（无效）的情况。 缺省情况下，失效链路定时器的值为 10s，可用 **undo mpls te cspf timer failed-link** 命令恢复缺省值

6.2.8　配置带宽的泛洪阈值

带宽的泛洪阈值是指 TE 隧道占用或释放的链路带宽与 TEDB 中剩余的链路带宽的比值，即链路带宽的变化值。当链路带宽变化很小时，每次变化都进行带宽泛洪会浪费网络资源。例如某条链路带宽为 100Mbit/s，在此链路上依次建立 100 条 1Mbit/s 的 TE 隧道时，则需要进行 100 次泛洪。

但如果设置一个泛洪阈值，则可以大大减少这样的带宽泛洪。如设置泛洪阈值为 10%，则建立第 1～9 条时不进行泛洪，当建立第 10 条时才对第 1～10 条所占用的 10Mbit/s 带宽进行泛洪。当建立第 11～18 条隧道时不进行泛洪，当建立第 19 条时才泛洪。依此类推。因此配置带宽泛洪阈值可减少泛洪次数，节约网络资源。

可在 MPLS TE 隧道入节点或中间节点上按照表 6-15 中步骤配置带宽的泛洪阈值，以减少带宽泛洪的次数，以降低因此而造成对网络性能的影响。

表 6-15　　配置带宽的泛洪阈值的步骤

步骤	命令	说明
1	**system-view** 例如：<Huawei> **system-view**	进入系统视图
2	**interface** *interface-type interface-number* 例如：[Huawei] **interface** gigabitethernet 1/0/0	进入 MPLS TE 链路的接口视图
3	**mpls te bandwidth change thresholds** { **down** | **Up** } *percent* 例如：[Huawei-GigabitEthernet1/0/0] **mpls te bandwidth change thresholds down** 10	配置带宽的泛洪阈值。命令中的参数和选项说明如下。 • **down**：二选一选项，指定 MPLS TE 隧道占用带宽的泛洪阈值。当 MPLS TE 隧道占用的带宽与 TEDB 中的链路剩余带宽的比值大于或等于此阈值时，IGP 将进行泛洪，CSPF 更新 TEDB

（续表）

步骤	命令	说明
3	**mpls te bandwidth change thresholds** { **down** | **Up** } *percent* 例如：[Huawei-GigabitEthernet1/0/0] **mpls te bandwidth change thresholds down** 10	• **Up**：二选一选项，MPLS TE 隧道释放带宽的泛洪阈值。当 MPLS TE 隧道释放的带宽与 TEDB 中的链路剩余带宽的比值大于或等于此阈值时，IGP 将进行泛洪，CSPF 更新 TEDB。 • *percent*：带宽阈值百分比，整数形式，取值范围是 0～100。 缺省情况下，当一条链路上为 MPLS TE 隧道保留的带宽与 TEDB 中的链路剩余带宽的比值等于或大于 10%，或 MPLS TE 隧道释放的带宽与 TEDB 中剩余带宽的比值等于或大于 10%时，IGP 将对该链路信息进行泛洪，CSPF 更新 TEDB，可用 **undo mpls te bandwidth change thresholds** { **down** | **Up** }命令恢复缺省设置

6.2.9　调整 CR-LSP 路径选择的配置管理

完成以上所需调整的 CR-LSP 路径选择的相关配置后，可执行以下 display 命令检查配置，验证配置结果。

■ **display mpls te tunnel verbose**：查看 MPLS TE 隧道信息。

■ **display mpls te srlg** { *srlg-number* | **all** }：查看 SRLG 配置和 SRLG 成员接口信息。

■ **display mpls te link-administration srlg-information** [**interface** *interface-type interface-number*]：查看接口所属的 SRLG。

■ **display mpls te tunnel c-hop** [*tunnel-name*] [**lsp-id** *ingress-lsr-id session-id lsp-id*]：查看隧道算路结果。

■ **display default-parameter mpls te cspf**：查看 CSPF 的缺省配置。

6.2.10　MPLS TE 隧道属性配置示例

如图 6-3 所示，LSRA 上要建立两条到达 LSRD 的动态 MPLS TE 隧道：Tunnel0/0/1 和 Tunnel0/0/2。现要求根据各链路上的管理组属性，使用隧道的亲和属性及掩码，使 LSRA 上的 Tunnel0/0/1 使用 LSRB→LSRC 的物理链路，Tunnel0/0/2 使用 LSRB→LSRE→LSRC 的物理链路。

1．基本配置思路分析

本示例的关键配置是各出接口的管理组属性，以及入节点 Tunnel 接口的亲和属性、掩码的匹配配置。总体原则是：如果希望某条链路能够被某条 CR-LSP 所用，则在入节点上配置的掩码中所有为 1 的位中，该链路的管理组属性中至少有 1 位与入节点上配置的亲和属性中的相应位都为 1；而且入节点上配置的亲和属性为 0 的位，该链路对应的管理组属性位不能为 1。

本示例要通过 MPLS TE 隧道属性来调整两条 CR-LSP 路径的选择，先要在入节点 LSRA 的对应 Tunnel 接口上配置亲和属性、掩码，然后根据以上匹配规则在 LSRA、LSRB、LSRC 和 LSRE 节点的出接口上配置对应的链路管理组属性。其中 LSRA 和 LSRC 的出接口要同时建立两条隧道的 CR-LSP，所以配置的链路属性是一致的，且所配置的管理

组属性通过与入节点亲和属性、掩码进行匹配时最终能同时被两条 CR-LSP 选择；LSRB 的两个出接口通过管理组属性的配置要分别被两条 CR-LSP 所选择；LSRE 的出接口通过管理组属性配置要被 Tunnel0/0/2 对应的隧道所选择。

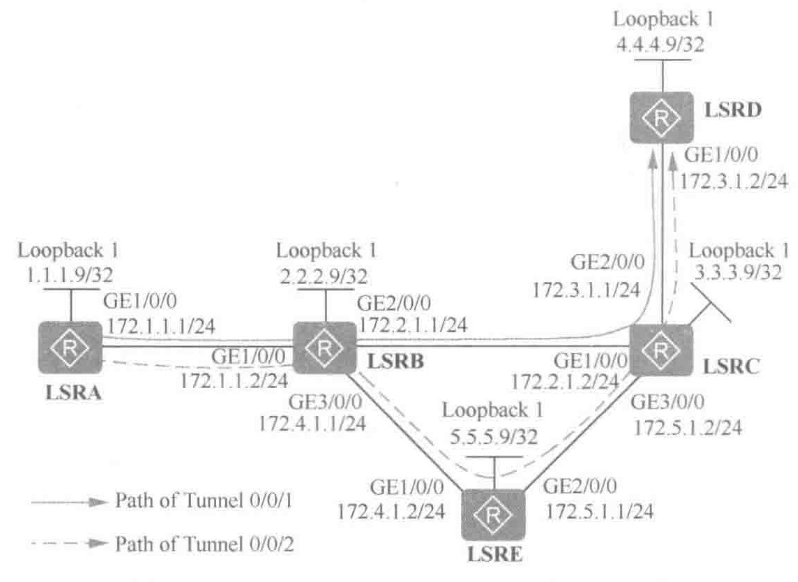


图 6-3 MPLS TE 隧道属性配置示例的拓扑结构

当然在配置 MPLS TE 隧道属性前仍然要先完成两条 CR-LSP 建立的配置。故本示例的基本配置思路如下。

（1）配置各节点的各接口（包括 Loopback 接口）IP 地址和 OSPF 协议，实现各节点之间公网路由可达。

（2）配置各节点的 LSR ID，并使能全局及各公网接口的 MPLS、MPLS TET 和 RSVP-TE 能力，在入节点上还要使能 CSPF 能力。本示例仅介绍由 LSRA 到 LSRD 的单向 TE 隧道配置，故此处的入节点仅为 LSRA。在实际通信中，要配置双向 TE 隧道。

（3）在各节点上配置 OSPF TE，使能 OSPF 的 MPLS TE 信息发布能力。

（4）在隧道入节点创建隧道接口，配置隧道的 IP 地址、隧道协议、目的地址、隧道 ID、动态信令协议 RSVP-TE、亲和属性和掩码。

（5）在各节点的隧道出接口（LSRD 上没有隧道出接口，故不需要配置）上配置链路的管理组属性，使这些出接口的链路管理组属性与对应 Tunnel 接口中上配置的亲和属性匹配后被对应的 CR-LSP 选择。

2. 具体配置步骤

（1）配置各节点的各接口的 IP 地址和 OSPF 协议，实现公网三层互通。

\# LSRA 上的配置。

```
<Huawei> system-view
[Huawei] sysname LSRA
[LSRA] interface gigabitethernet 1/0/0
[LSRA-GigabitEthernet1/0/0] ip address 172.1.1.1 255.255.255.0
[LSRA-GigabitEthernet1/0/0] quit
[LSRA] interface loopback 1
[LSRA-LoopBack1] ip address 1.1.1.9 255.255.255.255
[LSRA-LoopBack1] quit
```

```
[LSRA] ospf 1
[LSRA-ospf-1] area 0
[LSRA-ospf-1-area-0.0.0.0] network 1.1.1.9 0.0.0.0
[LSRA-ospf-1-area-0.0.0.0] network 172.1.1.0 0.0.0.255
[LSRA-ospf-1-area-0.0.0.0] quit
[LSRA-ospf-1] quit
```

#　LSRB 上的配置。

```
<Huawei> system-view
[Huawei] sysname LSRB
[LSRB] interface gigabitethernet 1/0/0
[LSRB-GigabitEthernet1/0/0] ip address 172.1.1.2 255.255.255.0
[LSRB-GigabitEthernet1/0/0] quit
[LSRB] interface gigabitethernet 2/0/0
[LSRB-GigabitEthernet2/0/0] ip address 172.2.1.1 255.255.255.0
[LSRB-GigabitEthernet2/0/0] quit
[LSRB] interface gigabitethernet 3/0/0
[LSRB-GigabitEthernet3/0/0] ip address 172.4.1.1 255.255.255.0
[LSRB-GigabitEthernet3/0/0] quit
[LSRB] interface loopback 1
[LSRB-LoopBack1] ip address 2.2.2.9 255.255.255.255
[LSRB-LoopBack1] quit
[LSRB] ospf 1
[LSRB-ospf-1] area 0
[LSRB-ospf-1-area-0.0.0.0] network 2.2.2.9 0.0.0.0
[LSRB-ospf-1-area-0.0.0.0] network 172.1.1.0 0.0.0.255
[LSRB-ospf-1-area-0.0.0.0] network 172.2.1.0 0.0.0.255
[LSRB-ospf-1-area-0.0.0.0] network 172.4.1.0 0.0.0.255
[LSRB-ospf-1-area-0.0.0.0] quit
[LSRB-ospf-1] quit
```

#　LSRC 上的配置。

```
<Huawei> system-view
[Huawei] sysname LSRC
[LSRC] interface gigabitethernet 1/0/0
[LSRC-GigabitEthernet1/0/0] ip address 172.2.1.2 255.255.255.0
[LSRC-GigabitEthernet1/0/0] quit
[LSRC] interface gigabitethernet 2/0/0
[LSRC-GigabitEthernet2/0/0] ip address 172.3.1.1 255.255.255.0
[LSRC-GigabitEthernet2/0/0] quit
[LSRC] interface gigabitethernet 3/0/0
[LSRC-GigabitEthernet3/0/0] ip address 172.5.1.2 255.255.255.0
[LSRC-GigabitEthernet3/0/0] quit
[LSRC] interface loopback 1
[LSRC-LoopBack1] ip address 3.3.3.9 255.255.255.255
[LSRC-LoopBack1] quit
[LSRC] ospf 1
[LSRC-ospf-1] area 0
[LSRC-ospf-1-area-0.0.0.0] network 3.3.3.9 0.0.0.0
[LSRC-ospf-1-area-0.0.0.0] network 172.2.1.0 0.0.0.255
[LSRC-ospf-1-area-0.0.0.0] network 172.3.1.0 0.0.0.255
[LSRC-ospf-1-area-0.0.0.0] network 172.5.1.0 0.0.0.255
[LSRC-ospf-1-area-0.0.0.0] quit
[LSRC-ospf-1] quit
```

#　LSRD 上的配置。

```
<Huawei> system-view
[Huawei] sysname LSRD
```

```
[LSRD] interface gigabitethernet 1/0/0
[LSRD-GigabitEthernet1/0/0] ip address 172.3.1.2 255.255.255.0
[LSRD-GigabitEthernet1/0/0] quit
[LSRD] interface loopback 1
[LSRD-LoopBack1] ip address 4.4.4.9 255.255.255.255
[LSRD-LoopBack1] quit
[LSRD] ospf 1
[LSRD-ospf-1] area 0
[LSRD-ospf-1-area-0.0.0.0] network 4.4.4.9 0.0.0.0
[LSRD-ospf-1-area-0.0.0.0] network 172.3.1.0 0.0.0.255
[LSRD-ospf-1-area-0.0.0.0] quit
[LSRD-ospf-1] quit
```

LSRE 上的配置。

```
<Huawei> system-view
[Huawei] sysname LSRE
[LSRE] interface gigabitethernet 1/0/0
[LSRE-GigabitEthernet1/0/0] ip address 172.4.1.2 255.255.255.0
[LSRE-GigabitEthernet1/0/0] quit
[LSRE] interface gigabitethernet 2/0/0
[LSRE-GigabitEthernet2/0/0] ip address 172.5.1.1 255.255.255.0
[LSRE-GigabitEthernet2/0/0] quit
[LSRE] interface loopback 1
[LSRE-LoopBack1] ip address 5.5.5.9 255.255.255.255
[LSRE-LoopBack1] quit
[LSRE] ospf 1
[LSRE-ospf-1] area 0
[LSRE-ospf-1-area-0.0.0.0] network 5.5.5.9 0.0.0.0
[LSRE-ospf-1-area-0.0.0.0] network 172.4.1.0 0.0.0.255
[LSRE-ospf-1-area-0.0.0.0] network 172.5.1.0 0.0.0.255
[LSRE-ospf-1-area-0.0.0.0] quit
[LSRE-ospf-1] quit
```

以上配置完成后，在各节点上执行 **display ip routing-table** 命令，应可以看到相互之间都学到了到对方 Loopback1 的路由。以下是在 LSRA 上执行该命令的输出示例（参见输出信息中的粗体字部分）。

```
[LSRA] display ip routing-table
Route Flags: R - relay, D - download to fib
------------------------------------------------------------------------------
Routing Tables: Public
         Destinations : 16        Routes : 16

Destination/Mask    Proto   Pre  Cost      Flags NextHop         Interface

        1.1.1.9/32  Direct  0    0           D   127.0.0.1       LoopBack1
        2.2.2.9/32  OSPF    10   1           D   172.1.1.2       GigabitEthernet1/0/0
        3.3.3.9/32  OSPF    10   2           D   172.1.1.2       GigabitEthernet1/0/0
        4.4.4.9/32  OSPF    10   3           D   172.1.1.2       GigabitEthernet1/0/0
        5.5.5.9/32  OSPF    10   2           D   172.1.1.2       GigabitEthernet1/0/0
      127.0.0.0/8   Direct  0    0           D   127.0.0.1       InLoopBack0
      127.0.0.1/32  Direct  0    0           D   127.0.0.1       InLoopBack0
127.255.255.255/32  Direct  0    0           D   127.0.0.1       InLoopBack0
      172.1.1.0/24  Direct  0    0           D   172.1.1.1       GigabitEthernet1/0/0
      172.1.1.1/32  Direct  0    0           D   127.0.0.1       GigabitEthernet1/0/0
    172.1.1.255/32  Direct  0    0           D   127.0.0.1       GigabitEthernet1/0/0
```

```
        172.2.1.0/24  OSPF    10   2        D   172.1.1.2       GigabitEthernet1/0/0
        172.3.1.0/24  OSPF    10   3        D   172.1.1.2       GigabitEthernet1/0/0
        172.4.1.0/24  OSPF    10   2        D   172.1.1.2       GigabitEthernet1/0/0
        172.5.1.0/24  OSPF    10   3        D   172.1.1.2       GigabitEthernet1/0/0
  255.255.255.255/32  Direct  0    0        D   127.0.0.1       InLoopBack0
```

（2）在各节点配置 MPLS LSR ID，使能 MPS、MPLS TE、RSVP-TE，在 LSRA 上使能 CSPF。本示例仅以从 LARA 到 LSRD 的单向隧道配置为例进行介绍。

\# LSRA 上的配置。

```
[LSRA] mpls lsr-id 1.1.1.9
[LSRA] mpls
[LSRA-mpls] mpls te
[LSRA-mpls] mpls rsvp-te
[LSRA-mpls] mpls te cspf
[LSRA-mpls] quit
[LSRA] interface gigabitethernet 1/0/0
[LSRA-GigabitEthernet1/0/0] mpls
[LSRA-GigabitEthernet1/0/0] mpls te
[LSRA-GigabitEthernet1/0/0] mpls rsvp-te
[LSRA-GigabitEthernet1/0/0] quit
```

\# LSRB 上的配置。

```
[LSRB] mpls lsr-id 2.2.2.9
[LSRB] mpls
[LSRB-mpls] mpls te
[LSRB-mpls] mpls rsvp-te
[LSRB-mpls] quit
[LSRB] interface gigabitethernet 1/0/0
[LSRB-GigabitEthernet1/0/0] mpls
[LSRB-GigabitEthernet1/0/0] mpls te
[LSRB-GigabitEthernet1/0/0] mpls rsvp-te
[LSRB-GigabitEthernet1/0/0] quit
[LSRB] interface gigabitethernet 2/0/0
[LSRB-GigabitEthernet2/0/0] mpls
[LSRB-GigabitEthernet2/0/0] mpls te
[LSRB-GigabitEthernet2/0/0] mpls rsvp-te
[LSRB-GigabitEthernet2/0/0] quit
[LSRB] interface gigabitethernet 3/0/0
[LSRB-GigabitEthernet3/0/0] mpls
[LSRB-GigabitEthernet3/0/0] mpls te
[LSRB-GigabitEthernet3/0/0] mpls rsvp-te
[LSRB-GigabitEthernet3/0/0] quit
```

\# LSRC 上的配置。

```
[LSRC] mpls lsr-id 3.3.3.9
[LSRC] mpls
[LSRC-mpls] mpls te
[LSRC-mpls] mpls rsvp-te
[LSRC-mpls] quit
[LSRC] interface gigabitethernet 1/0/0
[LSRC-GigabitEthernet1/0/0] mpls
[LSRC-GigabitEthernet1/0/0] mpls te
[LSRC-GigabitEthernet1/0/0] mpls rsvp-te
[LSRC-GigabitEthernet1/0/0] quit
[LSRC] interface gigabitethernet 2/0/0
[LSRC-GigabitEthernet2/0/0] mpls
```

```
[LSRC-GigabitEthernet2/0/0] mpls te
[LSRC-GigabitEthernet2/0/0] mpls rsvp-te
[LSRC-GigabitEthernet2/0/0] quit
[LSRC] interface gigabitethernet 3/0/0
[LSRC-GigabitEthernet3/0/0] mpls
[LSRC-GigabitEthernet3/0/0] mpls te
[LSRC-GigabitEthernet3/0/0] mpls rsvp-te
[LSRC-GigabitEthernet3/0/0] quit
```

LSRD 上的配置。

```
[LSRD] mpls lsr-id 4.4.4.9
[LSRD] mpls
[LSRD-mpls] mpls te
[LSRD-mpls] mpls rsvp-te
[LSRD-mpls] quit
[LSRD] interface gigabitethernet 1/0/0
[LSRD-GigabitEthernet1/0/0] mpls
[LSRD-GigabitEthernet1/0/0] mpls te
[LSRD-GigabitEthernet1/0/0] mpls rsvp-te
[LSRD-GigabitEthernet1/0/0] quit
```

LSRE 上的配置。

```
[LSRE] mpls lsr-id 5.5.5.9
[LSRE] mpls
[LSRE-mpls] mpls te
[LSRE-mpls] mpls rsvp-te
[LSRE-mpls] quit
[LSRE] interface gigabitethernet 1/0/0
[LSRE-GigabitEthernet1/0/0] mpls
[LSRE-GigabitEthernet1/0/0] mpls te
[LSRE-GigabitEthernet1/0/0] mpls rsvp-te
[LSRE-GigabitEthernet1/0/0] quit
[LSRE] interface gigabitethernet 2/0/0
[LSRE-GigabitEthernet2/0/0] mpls
[LSRE-GigabitEthernet2/0/0] mpls te
[LSRE-GigabitEthernet2/0/0] mpls rsvp-te
[LSRE-GigabitEthernet2/0/0] quit
```

（3）在各节点上配置 OSPF TE，使能 OSPF 的 MPLS 信息发布能力。

LSRA 上的配置。

```
[LSRA] ospf
[LSRA-ospf-1] opaque-capability enable
[LSRA-ospf-1] area 0
[LSRA-ospf-1-area-0.0.0.0] mpls-te enable
[LSRA-ospf-1-area-0.0.0.0] quit
[LSRA-ospf-1] quit
```

LSRB 上的配置。

```
[LSRB] ospf
[LSRB-ospf-1] opaque-capability enable
[LSRB-ospf-1] area 0
[LSRB-ospf-1-area-0.0.0.0] mpls-te enable
[LSRB-ospf-1-area-0.0.0.0] quit
[LSRB-ospf-1] quit
```

LSRC 上的配置。

```
[LSRC] ospf
[LSRC-ospf-1] opaque-capability enable
```

```
[LSRC-ospf-1] area 0
[LSRC-ospf-1-area-0.0.0.0] mpls-te enable
[LSRC-ospf-1-area-0.0.0.0] quit
[LSRC-ospf-1] quit
```

LSRD 上的配置。

```
[LSRD] ospf
[LSRD-ospf-1] opaque-capability enable
[LSRD-ospf-1] area 0
[LSRD-ospf-1-area-0.0.0.0] mpls-te enable
[LSRD-ospf-1-area-0.0.0.0] quit
[LSRD-ospf-1] quit
```

LSRE 上的配置。

```
[LSRE] ospf
[LSRE-ospf-1] opaque-capability enable
[LSRE-ospf-1] area 0
[LSRE-ospf-1-area-0.0.0.0] mpls-te enable
[LSRE-ospf-1-area-0.0.0.0] quit
[LSRE-ospf-1] quit
```

（4）在入节点创建并配置 MPLS TE Tunnel，其中关键的是两个 Tunnel 接口的亲和属性和掩码配置。两条隧道的亲和属性/掩码值对中至少有一个值不一样，且任何一个值均不能全为 0。

在 LSRA 上创建并配置 Tunnel0/0/1。

在 LSRA 的 Tunnel0/0/1 接口对应的隧道中，假设配置的亲和属性为 0x10101，掩码为 0x11011。

```
[LSRA] interface tunnel 0/0/1
[LSRA-Tunnel0/0/1] ip address unnumbered interface loopback 1
[LSRA-Tunnel0/0/1] tunnel-protocol mpls te
[LSRA-Tunnel0/0/1] destination 4.4.4.9
[LSRA-Tunnel0/0/1] mpls te tunnel-id 100
[LSRA-Tunnel0/0/1] mpls te record-route label
[LSRA-Tunnel0/0/1] mpls te affinity property 10101 mask 11011
[LSRA-Tunnel0/0/1] mpls te commit
[LSRA-Tunnel0/0/1] quit
```

在 LSRA 上创建 Tunnel0/0/2。

在 LSRA 的 Tunnel0/0/2 接口对应的隧道中，假设配置的亲和属性为 0x10001，掩码为 0x11101。

```
[LSRA] interface tunnel 0/0/2
[LSRA-Tunnel0/0/2] ip address unnumbered interface loopback 1
[LSRA-Tunnel0/0/2] tunnel-protocol mpls te
[LSRA-Tunnel0/0/2] destination 4.4.4.9
[LSRA-Tunnel0/0/2] mpls te tunnel-id 101
[LSRA-Tunnel0/0/2] mpls te record-route label
[LSRA-Tunnel0/0/2] mpls te affinity property 10011 mask 11101
[LSRA-Tunnel0/0/2] mpls te commit
[LSRA-Tunnel0/0/2] quit
```

（5）配置各节点出接口的管理组属性。因为本示例仅介绍从 LSRA 到 LSRD 的单方向 CR-LSP 配置，所以作为出节点的 LSRD 上不用配置管理组属性。

在入节点 LSRA 上配置好了隧道的亲和属性、掩码后，如果要使某链路被该隧道选中，则该链路管理组属性就不能随便配置了，必须满足以下两条基本原则。

■ 掩码中所有为 1 的位中，该链路的管理组属性中至少有 1 位与入节点上配置的亲和属性中的相应位都为 1。

■ 亲和属性为 0 的位，该链路对应的管理组属性位不能为 1。

其实，最终就是要使得亲和属性和掩码的逻辑“与”运算结果，与链路管理组属性与“掩码”的逻辑“与”运算结果一致，且结果不能全为 0。

LSRA 和 LSRC 上的配置。

前面已在 LSRA 上为 Tunnel0/0/1 和 Tunnel0/0/2 对应的隧道分别配置的亲和属性为 0x10101、0x10011，对应的掩码分别为 0x11011、0x11101。为简单起见，我们把以上配置的亲和属性和掩码以二进制看待（实际上是十六进制，每 1 位代表 4 位二进制）。由此可得出，Tunnel0/0/1、Tunnel0/0/2 的亲和属性与掩码的逻辑“与”运算结果都为 0x10001。

由于 LSRA 的 GE1/0/0 和 LSRC 的 GE20/0/0 接口是两条隧道的共同出接口，所以只需要配置一个链路管理组属性，即要能同时匹配两条隧道上配置的亲和属性。我们假设要配置的链路管理组属性为 xxxxx（因为前面在入节点上配置的新和属性和掩码都为 5 位十六进制，所以此处配置的链路管理组属性也为 5 位十六进制），然后分别与掩码 11011、11101 进行逻辑“与”运算（如图 6-4 所示），结果都要为 10001。根据逻辑“与”运算规则（在逻辑“与”运算中，相与的两位中只要有一位为 0，结果就为 0，两位都为 1 时才为 1）可以分析得出，LSRA 的 GE1/0/0 和 LSRC 的 GE20/0/0 接口的链路管理组属性只能是 10001。

```
  x x x x x            x x x x x
∧ 1 1 0 1 1          ∧ 1 1 1 0 1
-----------          -----------
  1 0 0 0 1            1 0 0 0 1
```

与 Tunnel0/0/1 隧道亲和属性掩码的逻辑“与”运算　　与 Tunnel0/0/2 隧道亲和属性掩码的逻辑“与”运算

图 6-4 LSRA 和 LSRC 出接口链路管理组属性与两隧道的亲和属性掩码的逻辑“与”运算

由此可得出在 LSRA 的 GE1/0/0 和 LSRC 的 GE20/0/0 接口链路管理组属性值为 10001。

```
[LSRA] interface gigabitethernet 1/0/0
[LSRA-GigabitEthernet1/0/0] mpls te link administrative group 10001
[LSRA-GigabitEthernet1/0/0] quit

[LSRC] interface gigabitethernet 2/0/0
[LSRC-GigabitEthernet2/0/0] mpls te link administrative group 10001
[LSRC-GigabitEthernet2/0/0] quit
```

LSRB 上的配置。

因为 LSRB 的两个出接口分别作为两条不同隧道的出接口，所以需要配置不同的管理组属性。假设在 LSRB 的 GE2/0/0 接口上配置的链路管理组属性值为 xxxxx，在 GE30/0/0 接口上配置的链路管理组属性值为 yyyyy，然后分别与两隧道上配置的掩码 11011、11101 进行逻辑“与”运算（如图 6-5 所示），结果也都要为 10001。

根据逻辑“与”运算规则可以得出 xxxxx 的值可以是 10001 或者 10101，而 yyyyy 的值可以是 10001 或 10011。而我们已知道，10001 这个是 LSRA 和 LSRC 出接口的链路管理属性，可以同时匹配两条隧道。但 LSRB 上的 GE2/0/0、GE30/0/0 接口均只需与一条隧道匹配，所要选择不同的链路管理属性，即要分别选择 10101 和 10011。

```
[LSRB] interface gigabitethernet 2/0/0
[LSRB-GigabitEthernet2/0/0] mpls te link administrative group 10101
[LSRB-GigabitEthernet2/0/0] quit
```

```
[LSRB] interface gigabitethernet 3/0/0
[LSRB-GigabitEthernet3/0/0] mpls te link administrative group 10011
[LSRB-GigabitEthernet3/0/0] quit
```

```
  x x x x x              y y y y y
∧ 1 1 0 1 1            ∧ 1 1 1 0 1
-----------            -----------
  1 0 0 0 1              1 0 0 0 1
```

GE2/0/0 接口链路管理属性与 Tunnel0/0/1 隧道亲和属性掩码的逻辑“与”运算

GE3/0/0 接口链路管理属性与 Tunnel0/0/2 隧道亲和属性掩码的逻辑“与”运算

图 6-5　LSRB 的两出接口分别与对应的隧道亲和属性掩码的逻辑“与”运算

LSRE 上的配置

LSRE 的出接口 GE2/0/0 也是作为 Tunnel0/0/2 接口对应隧道的出接口，所以它上面配置的管理属性与 LSRB 的 GE2/0/0 接口的管理组属性一样，也为 10011。

```
[LSRE] interface gigabitethernet 2/0/0
[LSRE-GigabitEthernet2/0/0] mpls te link administrative group 10011
[LSRE-GigabitEthernet2/0/0] quit
```

以上配置完成后，可在设备上查看 TEDB，其中包括各链路的 Color 字段，此即为各链路的管理组属性。以下是在 LSRA 上执行该命令的输出示例（参见输出信息中的粗体字部分）。

```
[LSRA] display mpls te cspf tedb node
 Router ID: 1.1.1.9
 IGP Type: OSPF          Process ID: 1
  MPLS-TE Link Count: 1
  Link[1]:
    OSPF Router ID: 1.1.1.9                    Opaque LSA ID: 1.0.0.1
    Interface IP Address: 172.1.1.1
    DR Address: 172.1.1.2
    IGP Area: 0
    Link   Type: Multi-access   Link Status: Active
    IGP Metric: 1                 TE Metric: 1              Color: 0x10001
...
```

3. 配置结果验证

以上配置完成后，执行 **display mpls te tunnel-interface** 命令在 LSRA 上查看 Tunnel 的状态，可以看到 Tunnel0/0/1 和 Tunnel0/0/2 状态为 Up。

```
[LSRA] display mpls te tunnel-interface
    ----------------------------------------------------------------
                              Tunnel0/0/1
    ----------------------------------------------------------------
    Tunnel State Desc    :  Up
    Active LSP           :  Primary LSP
    Session ID           :  100
    Ingress LSR ID       :  1.1.1.9          Egress LSR ID:  4.4.4.9
    Admin State          :  Up                Oper State   :  Up
    Primary LSP State    : Up
      Main LSP State     : READY                     LSP ID  : 47
```

```
--------------------------------------------------------------------------
                              Tunnel0/0/2
--------------------------------------------------------------------------
Tunnel State Desc     :  Up
Active LSP            :  Primary LSP
Session ID            :  101
Ingress LSR ID        :  1.1.1.9            Egress LSR ID    :  4.4.4.9
Admin State           :  Up                 Oper State       :  Up
Primary LSP State     :  Up
  Main LSP State      :  READY              LSP ID           :  4
```

执行 **display mpls te tunnel path** 命令查看隧道经过的路径，看最终的路径是否符合要求，也可验证所配置的隧道亲和属性、掩码及各链路的管理组属性是否正确。

```
[LSRA] display mpls te tunnel path
 Tunnel Interface Name : Tunnel0/0/1
 Lsp ID : 1.1.1.9 :100 :47
 Hop Information
  Hop 0      172.1.1.1
  Hop 1      172.1.1.2   Label 1065
  Hop 2      2.2.2.9   Label 1065
  Hop 3      172.2.1.1
  Hop 4      172.2.1.2   Label 1075
  Hop 5      3.3.3.9   Label 1075
  Hop 6      172.3.1.1
  Hop 7      172.3.1.2   Label 3
  Hop 8      4.4.4.9   Label 3

 Tunnel Interface Name : Tunnel0/0/2
 Lsp ID : 1.1.1.9 :101 :4
 Hop Information
  Hop 0      172.1.1.1
  Hop 1      172.1.1.2   Label 1067
  Hop 2      2.2.2.9   Label 1067
  Hop 3      172.4.1.1
  Hop 4      172.4.1.2   Label 1040
  Hop 5      5.5.5.9   Label 1040
  Hop 6      172.5.1.1
  Hop 7      172.5.1.2   Label 1077
  Hop 8      3.3.3.9   Label 1077
  Hop 9      172.3.1.1
  Hop 10     172.3.1.2   Label 3
  Hop 11     4.4.4.9   Label 3
```

6.3 调整 MPLS TE 隧道的建立

在建立 MPLS TE 隧道的过程中，根据实际应用还可能需要通过某些辅导配置来调整隧道的建立，以满足特定的需求，但一般不需要配置。

MPLS TE 提供多种方式，用于灵活调整 TE 隧道的建立，主要包括：环路检测、记录路由和标签、CR-LSP 重优化、隧道重建、RSVP 信令延迟触发功能和隧道的优先级。均为可选任务，无配置顺序限制，请根据实际情况进行配置。在调整 MPLS TE 隧道的建立之前，需要完成动态 MPLS TE 隧道或动态 DS-TE 隧道配置。

6.3.1 配置环路检测

MPLS TE 中的环路检测机制最大跳数为 32。当某一节点收到的路径信息记录表中已有该 LSR 的记录，或路径信息记录表中记录的路径跳数超过 32 跳时，均认为出现环路，所请求的 CR-LSP 会建立失败。通过配置环路检测功能，可以防止在 CR-LSP 建立时产生环路。

可在**入节点**的 Tunnel 接口视图下执行 **mpls te loop-detection** 命令配置隧道建立时进行环路检测，然后执行 **mpls te commit** 命令提交隧道当前配置，使配置更改生效。缺省情况下，不进行环路检测，可用 **undo mpls te loop-detection** 命令去使能环路检测功能。

6.3.2 配置记录路由和标签

在没有配置显式路径的情况下，MPLS TE Tunnel 建立成功后，系统不会记录隧道的详细路径。这样一来，如果需要查看隧道的详细路径就没办法了。此时可以在入节点的 Tunnel 接口下配置在 Path、Resv 消息中携带 RRO 对象，记录消息经过每一跳的 IP 地址，同时还可以选择记录经过每一跳的标签，这样就可以查看隧道的详细路径和每一跳所分配的标签分配信息了。

可在**入节点**的 Tunnel 接口视图下通过 **mpls te record-route [label]**命令使能建立隧道时记录路由和标签功能，如果不选择可选项 **label**，则不记录经过每一跳的标签，然后执行 **mpls te commit** 命令提交隧道当前配置，使配置更改生效。但如果网络规模比较大，不建议使用此命令，因为 RRO 中记录了每一跳的 IP 地址，如果跳数很多，则会导致 Path 或 Resv 消息过大，降低系统性能。缺省情况下，未使能建立隧道时记录路由和标签功能，可用 **undo mpls te record-route** 命令去使能建立隧道时记录路由和标签功能。

6.3.3 配置 CR-LSP 重优化

TE 隧道建立好后，如果网络拓扑结构，或者设备链路配置发生了改变，对于隧道两端的通信路径可能有更好的选择。此时可以通过配置隧道重优化功能，让系统自动定期重计算 CR-LSP 穿越的路由。如果发现重计算的路由优于当前路由，则创建一条新的 CR-LSP，并为之分配新路由，将业务从旧的 CR-LSP 切换至新的 CR-LSP，删除旧 CR-LSP。这样就可以经常保持隧道两端通信的最佳性能。

CR-LSP 重优化功能的配置步骤见表 6-16，也是仅可在**入节点**上配置。

表 6-16　　配置 CR-LSP 重优化的步骤

步骤	命令	说明
1	**system-view** 例如：<Huawei> **system-view**	进入系统视图
2	**interface tunnel** *interface-number* 例如：[Huawei] **interface tunnel** 0/0/1	进入 MPLS TE 隧道的 Tunnel 接口视图
3	**mpls te reoptimization** [**frequency** *interval*] 例如：[Huawei-Tunnel0/0/1] **mpls te reoptimization frequency** 43200	配置定时重优化。可选参数 **frequency***interval* 用来指定重优化频率，整数形式，取值范围是 60～604800，单位是秒，缺省值是 3600s。即每隔 *interval* 周期都会根据 TE 隧道的约束条件执行计算，如果有到达同一

（续表）

步骤	命令	说明
3	**mpls te reoptimization** [**frequency** *interval*] 例如：[Huawei-Tunnel0/0/1] **mpls te reoptimization frequency** 43200	目的地址的更优路径，则对 CR-LSP 进行重优化。 缺省情况下，不进行重优化，可用 **undo mpls te reoptimization** 命令恢复缺省配置
4	**mpls te commit** 例如：[Huawei-Tunnel0/0/1] **mpls te commit**	提交隧道当前配置。凡发生了 TE 隧道配置更改，均要执行本命令，才能使配置更改生效
5	**quit** 例如：[Huawei-Tunnel0/0/1] **quit**	返回系统视图
6	**mpls** 例如：[Huawei] **mpls**	进入 MPLS 视图
7	**mpls te switch-delay** *switch-time* **delete-delay** *delete-time* 例如：[Huawei-mpls] **mpls te switch-delay** 3000 **delete-delay** 8000	（可选）配置切换到新 CR-LSP 的延迟时间和删除旧 CR-LSP 的延迟时间。命令中的参数说明如下。 • **switch-delay** *switch-time*：指定 TE 流量从旧 CR-LSP 切换到新 CR-LSP 的延迟时间，整数形式，取值范围是 0～600000，单位是毫秒。 • **delete-delay** *delete-time*：指定 TE 流量切换到新 CR-LSP 后，删除旧 CR-LSP 的延迟时间，整数形式，取值范围是 0～600000，单位是毫秒。 缺省情况下，切换延迟时间为 5000ms，删除延迟时间为 7000ms，可用 **undo mpls te switch-delay** *switch-time* **delete-delay** *delete-time* 命令恢复缺省配置
8	**return** 例如：[Huawei-mpls] **return**	退回用户视图
9	**mpls te reoptimization** [**tunnel** *interface-number*] 例如：<Huawei> **mpls te reoptimization**	手动触发隧道的重优化功能。在隧道视图下通过本表第 3 步配置定时重优化后，可以通过本命令手动触发隧道重优化进程，当然也可以在到达重优化时间间隔后由系统自动触发重集成自动化。可选参数 **tunnel** *interface-number* 用来指定立即进行重优化对应的隧道，如果不指定此可选参数，则手动触发所有配置了定时重优化功能的隧道的重优化进程。 隧道重优化包括以下两种方式。 • 自动重优化：即系统周期性地对 TE 隧道进行重优化，无需人工干预，从而节省人力。系统通过本表第 3 步的配置即可实现此功能，但对当前 TE 隧道生效。 • 手工重优化：当用户需要立即对 TE 隧道进行重优化时，可以配置本命令对指定或所有配置了定时重优化功能的 TE 隧道进行手工重优化。 手工重优化功能主要应用于以下两种场景。 • 手工调整了网络拓扑结构，并且需要 TE 隧道立即选用调整后的最优路径。 • 需要批量地对 TE 隧道进行立即重优化，及时优化 TE 隧道的资源利用。 执行了手动重优化后，定时重优化的定时器将被清零，重新计时

6.3.4　配置隧道重建

Tunnle 接口建立之后，本地节点会定时向邻居节点发送 Path 消息，并接收邻居回复的 Resv 消息，来维持 CR-LSP 的 Up 状态。如果在规定的重建时间间隔内，没有收到邻居发来的 Resv 消息，则本地节点认为该 CR-LSP 进入 Down 状态，尝试重新建立 CR-LSP。如果建立不成功，系统会每隔一定时间（即配置的重建隧道的间隔）开始新一轮重建。

可以配置当 CR-LSP 建立不成功时，后续发起隧道重建的时间间隔，也可以实现与 6.3.3 节介绍的 CR-LSP 重优化功能类似的、定期重新计算 CR-LSP 穿越的路由目的。但是，本节所介绍的隧道重建仅是在当前 CR-LSP 状态为 Down 时所进行的 CR-LSP 重建，而不是针对当前 Up 状态的 CR-LSP 优化。新的 CR-LSP 创建后会为之分配新路由，并将业务从旧的 CR-LSP 切换至新的 CR-LSP，删除旧 CR-LSP。

可在隧道**入节点**的 Tunnel 接口视图下通过 **mpls te timer retry** *interval* 命令配置每轮发起重建隧道的间隔时间，整数形式，取值范围是 10～65535，单位为秒，使在当前 CR-LSP 状态为 Down 时发起隧道重建，然后执行 **mpls te commit** 命令提交隧道当前配置，使配置更改生效。缺省情况下，重建隧道的时间间隔为 30s，可用 **undo mpls te timer retry** 命令恢复缺省设置。

6.3.5　配置 RSVP 信令延迟触发功能和隧道优先级

当 MPLS 网络出现故障，需要重新创建大量 RSVP CR-LSP 时，重建大量 RSVP CR-LSP 需要占用不少系统资源。如果配置信令延时触发，则可以降低创建 RSVP CR-LSP 所占用的系统资源。

这需要在有大量 CR-LSP 经过的节点的 MPLS 视图下执行 **mpls te signaling-delay-trigger enable** 命令，使能 RSVP 信令延迟触发功能。缺省情况下，未使能 RSVP 信令延迟触发功能，可用 **undo mpls te signaling-delay-trigger enable** 命令去使能 RSVP 信令延时触发功能。

如果在建立 CR-LSP 的过程中，无法找到满足所需带宽要求的路径，可以根据建立优先级和保持优先级进行抢占。可在 MPLS TE 隧道入节点的 Tunnel 接口视图下通过 **mpls te priority** *setup-priority* [*hold-priority*]命令配置隧道的建立优先级和保持优先级。命令中的参数说明如下。

■ *setup-priority*：指定隧道的建立优先级，整数形式，取值范围是 0～7，数值越小则优先级越高。配置的建立优先级的值不应该小于保持优先级的值。

■ *hold-priority*：可选参数，指定隧道的保持优先级整数形式，取值范围是 0～7，数值越小则优先级越高。当不配置保持优先级时，保持优先级与建立优先级相同。

配置隧道优先级后，执行 **mpls te commit** 命令提交隧道当前配置，使配置更改生效。缺省情况下，建立优先级和保持优先级的值都为 7（最低优先级），可用 **undo mpls te priority** 命令恢复缺省设置。

当已经完成所有调整 MPLS TE 隧道建立功能的配置后，可通过 **display mpls te tunnel-interface** [**tunnel** *interface-number*]命令查看隧道接口信息。

第7章 MPLS TE可靠性功能配置与管理



第 6 章介绍了 MPLS TE 隧道在参数调整方面的配置与管理方法，本章再来介绍 MPLS TE 隧道在可靠性方面的功能配置与管理方法。

在 MPLS TE 隧道可靠性功能包括许多方面，如备份 CR-LSP、静态/动态 BFD 检测、TE FRR（快速重路由）、共享风险链路组、RVSP GR（平滑重启）等。由于篇幅限制，在此仅介绍最常用的备份 CR-LSP，以及静态/动态 BFD for CR-LSP 的配置与管理方，通过它们可以进一步提高 MPLS TE 隧道的可用性和可靠性。

7.1 CR-LSP 备份配置与管理

为了提高 MPLS TE 隧道的可靠性，可在 MPLS TE 隧道中配置备份 CR-LSP。配置了备份 CR-LSP 后，当入节点感知到主 CR-LSP 不可用时，会将流量切换到备份路径上，而当主 CR-LSP 路径恢复后又可将流量再切换回主 CR-LSP，以实现对主 CR-LSP 路径的备份保护。

本节将具体介绍 CR-LSP 备份实现原理，以及备份 CR-LSP 的配置与管理方法。

7.1.1 CR-LSP 备份实现原理

CR-LSP 备份除了有通常见到的“热备份”和“普通备份”两种方式外，为了进一步提高 MPLS TE 隧道的可靠性，系统还提供了一种“逃生路径”技术。下面先来了解这几种备份方式对应的概念。

■ 热备份（Hot-standby）：指在创建主 CR-LSP 后**立即**创建备份 CR-LSP。当主 CR-LSP 故障时，会自动将业务流量切换至备份 CR-LSP。

■ 普通备份：指在主 CR-LSP 故障后再创建备份 CR-LSP，将业务流量切换至备份 CR-LSP。它与热备份的区别仅体现在创建的时机，热备份是要创建主 CR-LSP 后自动随即创建的，而普通备份是仅当主 CR-LSP 出了故障后才按配置要求创建的。

■ 逃生路径：指在主、备 CR-LSP（可以是热备份 CR-LSP，也可以是普通备份 CR-LSP）都出现故障时，由系统根据配置自动创建一条临时的 CR-LSP，然后将业务流量切换到逃生路径上。但逃生路径没有带宽保证（只要能通即可），可以根据通过配置逃生路径的亲和属性和跳数限制来控制其途经的路径。

如图 7-1 所示，主 CR-LSP 路径为 PE1→P1→P2→PE2；备份 CR-LSP 路径为 PE1→P3→PE2。当主备 CR-LSP 都故障时，PE1 触发建立逃生路径 PE1→P4→PE2。

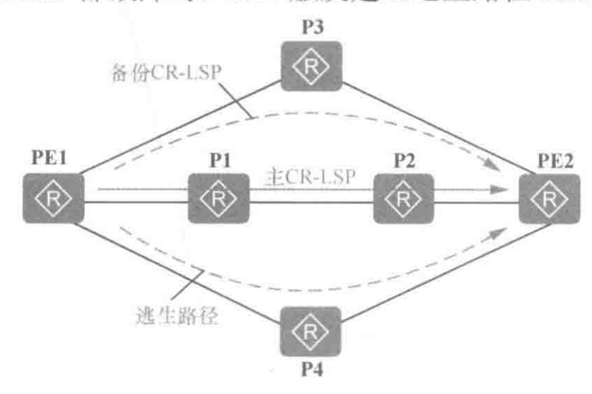


图 7-1 逃生路径示意

CR-LSP所支持的以上三种备份模式比较见表7-1。

表7-1　CR-LSP备份模式

备份模式	创建时机	优点	缺点
热备份模式	系统在建立主CR-LSP的同时，会创建一条与主CR-LSP路径分离的备份CR-LSP	流量切换速度快	如果没有配置热备份CR-LSP动态带宽功能，则需要占用额外的带宽
普通备份模式	当主CR-LSP失效后，系统将使用其他显式路径建立CR-LSP	无需占用主CR-LSP所在链路额外的带宽	流量切换速度不如热备份模式快
逃生路径模式	当主CR-LSP和备份CR-LSP都失效后，系统将使用剩余的路径建立逃生路径	对路径建立的要求较为宽松，容易建立	可能会降低某些QoS保证

CR-LSP备份技术的整个实现可以分为如下几个过程：①CR-LSP备份部署规划→②创建备份CR-LSP→③修改备份CR-LSP属性→④故障检测→⑤流量正切→⑥流量回切。下面分别予以介绍。

1. CR-LSP备份部署规划

部署CR-LSP备份时的路径和带宽规划所需考虑的事项见表7-2。

表7-2　CR-LSP备份的部署规划

部署子项	热备份	普通备份	逃生路径
路径	可以通过配置指明是否允许主/备路径部分重合。当备份CR-LSP使能显式路径建立时，以显式路径作为约束条件建立备份CR-LSP。 热备份CR-LSP支持如下约束条件。 • 显式路径。 • 亲和属性。 • 跳数限制。 • Overlap-path功能（具体参见下面的说明①）	无论备份CR-LSP是否用显式路径建立，备份CR-LSP的创建路径都可以与主CR-LSP的路径部分重合。 普通备份CR-LSP支持如下约束条件。 • 显式路径。 • 亲和属性。 • 跳数限制	由隧道入节点自动计算得出。 逃生路径支持如下约束条件。 • 亲和属性。 • 跳数限制
带宽	默认情况热备份CR-LSP带宽与主CR-LSP带宽值相等。支持dynamic-bandwidth技术后，可以保证热备份CR-LSP不承载流量时不会额外占用带宽（具体参见下面的说明②）	带宽始终与主CR-LSP带宽值相等	逃生路径不会在路径上预留带宽，只具有路径保护能力
配置组合	可以与逃生路径同时部署，共同保护主CR-LSP	只能单独作为备份路径保护主CR-LSP	—

说明

① 热备份CR-LSP可以配置Overlap-path功能，即在满足热备份CR-LSP的路径与主CR-LSP的路径尽量分离的情况下，支持部分重合，从而保证热备份CR-LSP对主CR-LSP的保护。

② 热备份CR-LSP还可以配置dynamic-bandwidth保护，即动态带宽保护功能。在该功能下，在主CR-LSP出现故障之前，热备份CR-LSP并不会额外占用网络中的带宽资源（带宽值为0），只有当热备份CR-LSP真正承载流量后才会占用网络的带宽资源。

这样可以更大幅度的节省网络资源，缩减网络开销。具体过程如下。

(1) 当主 CR-LSP 出现故障后，流量立即切换到带宽值为 0 的热备份 CR-LSP，同时 MPLS TE 隧道入节点立即采用 Make-Before-Break 机制重建热备份 CR-LSP。

(2) 当新的热备份 CR-LSP 建立成功后，流量就切换到新的热备份 CR-LSP 上，同时将最初的 0 带宽热备份 CR-LSP 删除。

(3) 当主 CR-LSP 故障恢复后，流量会重新回切到主 CR-LSP。此时热备份 CR-LSP 会释放已占用的带宽，重新采用 0 带宽建立热备份 CR-LSP。

2. 创建备份 CR-LSP

同一条隧道下可能存在表 7-1 所示的多种建立备份 CR-LSP 的方式。当新提交一条隧道或者隧道状态变为 Down 时，系统将按一定的优先级顺序轮流尝试创建热备份 CR-LSP、普通备份 CR-LSP、逃生路径，直到隧道建立成功。

3. 修改备份 CR-LSP 属性

当用户修改了备份 CR-LSP 的约束条件时，入节点会采用 Make-Before-Break 机制触发重新创建备份 CR-LSP。当携带新属性的备份 CR-LSP 完全建立成功以后，如果此时原备份 CR-LSP 已经承载了流量，MPLS TE 隧道会将流量切换到新的备份 CR-LSP 上，然后删除原备份 CR-LSP。

4. 故障检测

CR-LSP 备份技术可以采用如下故障检测技术。

■ RSVP-TE 的默认错误通告机制，但通常检测速度稍慢。

■ BFD for CR-LSP：可以对故障进行快速检测，推荐采用此种方式，具体配置方法将在本章 7.2 节介绍。

5. 流量正切

当隧道主 CR-LSP 发生故障后，入节点会触发流量从主 CR-LSP 向备份 CR-LSP 切换。其中切换的优先级顺序为：热备份优先级最高，其次是普通备份，逃生路径优先级最低。

6. 流量回切

在备份 CR-LSP 承载流量的期间，流量会根据具体情况，总是会试图按照一定优先级进行路径回切。其中主 CR-LSP 具有最高优先级，其次是热备份 CR-LSP，普通备份 CR-LSP 优先级最低。

7.1.2 CR-LSP 备份配置任务

配置 CR-LSP 备份需要在设备上进行以下配置，其中配置流量强制切换、配置热备份 CR-LSP 动态带宽功能和配置逃生路径为可选步骤。

(1) 创建备份 CR-LSP，可以选择创建的 CR-LSP 备份模式。

如果选择了 CR-LSP 热备份模式，为了实现毫秒级的快速切换，需要同时配置静态 BFD for CR-LSP 或者配置动态 BFD for CR-LSP。具体将在 7.2 节介绍。

(2)(可选) 配置流量强制切换。

(3)(可选) 配置热备份 CR-LSP 动态带宽保护功能。

(4)(可选) 配置逃生路径

配置 CR-LSP 备份之前，需要完成以下任务。

■ 配置动态 MPLS TE 隧道或者配置动态 DS-TE 隧道。
■ 在备份 CR-LSP 各节点的全局和接口下使能 MPLS、MPLS TE 和 RSVP-TE。

7.1.3 创建备份 CR-LSP

主隧道配置 CR-LSP 备份后，当主 CR-LSP 故障时，流量会切换到备份 CR-LSP 上，从而提供了端到端的保护。可在隧道**入节点**上按表 7-3 所示步骤创建备份 CR-LSP。

表 7-3　创建备份 **CR-LSP** 的步骤

步骤	命令	说明
1	**system-view** 例如：<Huawei> **system-view**	进入系统视图
2	**interface tunnel** *tunnel-number* 例如：[Huawei] **interface tunnel** 0/0/1	进入 MPLS TE 隧道的 Tunnel 接口视图
3	**mpls te backup** { **hot-standby** | **ordinary** } 例如：[Huawei-Tunnel0/0/1] **mpls te backup hot-standby**	配置当前隧道使用的备份模式，命令中的选项说明如下。 • **hot-standby**：二选一选项，配置热备份模式，主 CR-LSP 和备份 CR-LSP 同时存在，主 CR-LSP 失效时，直接将业务切换至备份 CR-LSP。 • **ordinary**：二选一选项，配置普通备份模式，主 CR-LSP 失效后才创建备份 CR-LSP。 配置热备份或普通备份模式后，系统会自动选择备份 CR-LSP 的路径。如果用户希望流量沿着指定的备份 CR-LSP 通过，可以继续执行表中第 4～6 步中的一个或多个步骤；当配置热备份模式时，还可以额外选配表中第 7～9 步中的一个或多个步骤。 【注意】同一个 Tunnel 接口不能同时作为备份 CR-LSP 和旁路隧道，既不能同时配置 **mpls te backup** 命令和 **mpls te bypass-tunnel** 命令，也不能同时配置 **mpls te backup** 命令和 **mpls te protected-interface** 命令（指定旁路隧道所要保护的接口）。 缺省情况下，隧道不进行备份，可用 **undo mpls te backup** { **hot-standby** | **ordinary** }恢复缺省配置
4	**mpls te path explicit-path** *path-name* **secondary** 例如：[Huawei-Tunnel0/0/1] **mpls te path explicit-path** path1	（可选）指定备份 CR-LSP 使用的显式路径。在配置本命令之前，必须先在系统视图下执行 **explicit-path** *path-name* 命令配置相应的显式路径，并通过 **next hop** 命令为显式路径指定节点，否则该命令配置不成功。 【注意】在规划备份 CR-LSP 使用的显式路径时，注意不要与主 CR-LSP 完全重合，否则无法达到保护的目的。 缺省情况下，没有为备份 CR-LSP 配置显式路径，可用 **undo mpls te path explicit-path** *path-name* secondary 命令删除原来的指定显式路径
5	**mpls te affinity property** *properties* [**mask** *mask-value*] **secondary** 例如：[Huawei-Tunnel0/0/1] **mpls te affinity property** a04 **mask** e0c **secondary**	（可选）配置备份 CR-LSP 的亲和属性。命令中的参数说明如下。 • *properties*：指定备份 CR-LSP 使用链路的亲和属性值，十六进制形式，取值范围是 0x0～0xFFFFFFFF，每一位代表一种属性

（续表）

步骤	命令	说明
5	**mpls te affinity property** *properties* [**mask** *mask-value*] **secondary** 例如：[Huawei-Tunnel0/0/1] **mpls te affinity property** a04 **mask** e0c **secondary**	• *mask-value*：可选参数，指定备份 CR-LSP 亲和属性掩码，即需要检查的链路管理组属性位，十六进制形式，取值范围是 0x0～0xFFFFFFFF，每一位代表一种属性。 【说明】亲和属性要与链路管理组属性配合使用，用来决定隧道是否选择某出接口对应的路径。有关亲和属性和链路管理组属性的组合应用配置方法及示例将参见本书第 6 章。 缺省情况下，备份 CR-LSP 的亲和属性值为 0x0，掩码为 0x0，可用 **undo mpls te affinity property secondary** 命令恢复缺省值
6	**mpls te hop-limit** *hop-limit-value* **secondary** 例如：[Huawei-Tunnel0/0/1] **mpls te hop-limit** 10 **secondary**	（可选）限制该备份 CR-LSP 的路径跳数，整数形式，取值范围是 1～32。 缺省情况下，备份 CR-LSP 的路径跳数是 32，可用 **undo mpls te hop-limit secondary** 命令恢复缺省值
7	**mpls te backup hot-standby overlap-path** 例如：[Huawei-Tunnel0/0/1] **mpls te backup hot-standby overlap-path**	（可选）使能热备份路径可以与主 CR-LSP 的路径重合功能。配置该功能后，当热备份 CR-LSP 不能完全排除主 CR-LSP 的路径时，需要通过本命令的配置允许热备份 CR-LSP 的路径与主 CR-LSP 的路径部分重合。 缺省情况下，热备份 CR-LSP 的路径不可以与主 CR-LSP 的路径重合。如果网络拓扑结构不能满足这个条件，将会导致热备份 CR-LSP 建立失败。可用 **undo mpls te backup hot-standby overlap-path** 命令去使能热备份 CR-LSP 的路径可以与主 CR-LSP 的路径重合功能
8	**mpls te backup hot-standby wtr** *interval* 例如：[Huawei-Tunnel0/0/1] **mpls te backup hot-standby wtr** 100	（可选）配置热备份回切的时间，整数形式，取值范围是 0～2592000，单位是秒。 缺省情况下，热备份回切的时间为 10s，可用 **undo mpls te backup hot-standby wtr** 命令恢复缺省配置
9	**mpls te backup hot-standby mode** { **revertive** [**wtr** *interval*] | **non-revertive** } 例如：[Huawei-Tunnel0/0/1] **mpls te backup hot-standby mode non-revertive**	（可选）指定热备份回切模式，即当备份 CR-LSP 承载流量时，试图回切到主 CR-LSP，或其他备份 CR-LSP（包括热备份 CR-LSP 和普通备份 CR-LSP）。命令中的参数和选项说明如下。 • **revertive**：二选一选项，表示模式为回切。 • **non-revertive**：二选一选项，表示模式为非回切。 • **wtr** *interval* ：配置热备份回切的时间（即回切延时），整数形式，取值范围是 0～2592000，单位是秒。缺省值是 10s。 缺省情况下，模式为回切，可用 **undo mpls te backup hot-standby mode** 命令恢复缺省配置
10	**mpls te commit** 例如：[Huawei-Tunnel0/0/1] **mpls te commit**	提交隧道配置，使配置更改生效

7.1.4 配置流量强制切换

在备份 CR-LSP 创建成功后，**当需要对主 CR-LSP 路径进行调整时**，可将流量强制

切换到备份 CR-LSP 上；当主 CR-LSP 完成调整时，再将流量切换回主 CR-LSP。这就保证了在 CR-LSP 路径调整过程中业务流量不中断。

流量强制切换配置要分两步进行，在对主 CR-LSP 路径进行调整前，需在 MPLS TE 隧道入节点的 Tunnel 接口视图上执行 **hotstandby-switch force** 命令，将流量强制切换到备份 CR-LSP 上，当然必须确保备份 CR-LSP 建立成功，否则将导致流量丢失；而在完成主 CR-LSP 路径调整后，需要在 MPLS TE 隧道入节点的 Tunnel 接口视图上执行 **hotstandby-switch clear** 命令，将流量强制切换回主 CR-LSP 上。

7.1.5　配置热备份 CR-LSP 动态带宽保护功能

一般情况创建热备份 CR-LSP 需要占用额外的带宽资源，可配置热备份动态带宽保护功能，使得**在创建主隧道的同时创建带宽为零的热备份 CR-LSP**。配置了热备份动态带宽保护功能后，在主 CR-LSP 出现故障之前，热备份 CR-LSP 并不会额外占用网络中的带宽资源（带宽值为零），只有当热备份 CR-LSP 真正承载流量后才会占用网络的带宽资源。这样可以更大幅度地节省网络资源，缩减网络开销。

当主 CR-LSP 发生故障时，原来的零带宽热备份 CR-LSP 开始承载流量，同时系统将采用 Make-Before-Break 机制，再新建一条满足带宽要求的热备份 CR-LSP，成功后再将流量从最初创建的零带宽热备份 CR-LSP 切换到新的热备份 CR-LSP 上，然后将零带宽的热备份 CR-LSP 删除。但在新建热备份 CR-LSP 时如果带宽资源不足，则流量又会切换到最初创建的带宽为 0 的热备份 CR-LSP，确保切换过程中流量不中断。

配置热备份 CR-LSP 动态带宽保护功能的方法也很简单，只需在具体的 Tunnel 接口视图下执行 **mpls te backup hot-standby dynamic-bandwidth** 命令即可。如果已经使用热备份模式创建了一条热备份 CR-LSP，那么配置热备份 CR-LSP 动态带宽功能后，系统将采用 Make-Before-Break 机制重建一条带宽值为零的热备份 CR-LSP，并替代原来的热备份 CR-LSP。配置好后，要在 Tunnel 接口视图下执行 **mpls te commit** 命令，提交配置，以使配置更改生效。

缺省未启用热备份 CR-LSP 动态带宽保护功能，可用 **undo mpls te backup hot-standby dynamic-bandwidth** 命令去使能热备份 CR-LSP 动态带宽保护功能，使热备份 CR-LSP 重新占用带宽。

7.1.6　配置逃生路径

当主隧道的入节点配置逃生路径后，在主 CR-LSP 和备份 CR-LSP 都发生故障时，流量会切换到逃生路径上。从中可以看出，逃生路径是流量切换的最后备份选择。主隧道的逃生路径也是在入节点上配置的，具体的配置步骤见表 7-4。

表 7-4　逃生路径的配置步骤

步骤	命令	说明
1	**system-view** 例如：<Huawei> **system-view**	进入系统视图
2	**interface tunnel** *tunnel-number* 例如：[Huawei] **interface tunnel** 0/0/1	进入 MPLS TE 隧道的 Tunnel 接口视图

（续表）

步骤	命令	说明
3	**mpls te backup ordinary best-effort** 例如：[Huawei-Tunnel0/0/1] **mpls te backup ordinary best-effort**	配置逃生路径。如果用户希望逃生路径沿着指定的路径建立，可以选择配置下面的第 4、第 5 步。 【注意】逃生路径不能和手工配置的普通备份模式同时配置，即不能同时配置本命令和 **mpls te backup ordinary** 命令
4	**mpls te affinity property** *properties* [**mask** *mask-value*] **best-effort** 例如：[Huawei-Tunnel0/0/1] **mpls te affinity property** a04 **mask** e0c **best-effort**	（可选）配置逃生路径的亲和属性。命令中的参数参见 7.1.3 节表 7-3 的第 5 步。 缺省情况下，逃生路径的亲和属性值为 0x0，掩码为 0x0，可用 **undo mpls te affinity property** best-effort 命令恢复缺省值
5	**mpls te hop-limit** *hop-limit-value* **best-effort** 例如：[Huawei-Tunnel0/0/1] **mpls te hop-limit** 10 **best-effort**	（可选）限制该逃生路径的路径跳数，整数形式，取值范围是 1～32。 缺省情况下，逃生路径的路径跳数是 32，可用 **undo mpls te hop-limit best-effort** 命令恢复缺省值
6	**mpls te commit** 例如：[Huawei-Tunnel0/0/1] **mpls te commit**	提交隧道配置，使配置更改生效

7.1.7 CR-LSP 备份配置管理

已经完成 CR-LSP 备份功能的所有配置后，可执行以下 **display** 命令查看相关配置，验证配置结果。

■ **display mpls te tunnel-interface** [**tunnel** *tunnel-number*]：查看指定或所有隧道接口的相关配置信息。

■ **display mpls te hot-standby state** { **all** [**verbose**] | **interface tunnel** *interface-number* }：查看所有或指定隧道的热备份状态信息。

■ **display mpls te tunnel** [**destination** *ip-address*] [**lsp-id** *ingress-lsr-id session-id local-lsp-id*] [**lsr-role** { **all** | **egress** | **ingress** | **remote** | **transit** }] [**name** *tunnel-name*] [{ **incoming-interface** | **interface** | **outgoing-interface** } *interface-type interface-number*] [**te-class0** | **te-class1** | **te-class2** | **te-class3** | **te-class4** | **te-class5** | **te-class6** | **te-class7**] [**verbose**]：按指定条件查看相关隧道信息。

7.1.8 CR-LSP 热备份配置示例

在如图 7-2 所示的 MPLS VPN 网络中，要从 LSRA 上建立一条 TE 隧道，目的地址为 LSRC，并配置 CR-LSP 热备份和逃生路径。各路径所经过的节点如下。

■ 主 CR-LSP 的路径为 LSRA→LSRB→LSRC。

■ 热备份 CR-LSP 的路径为 LSRA→LSRD→LSRC。

■ 逃生路径为 LSRA→LSRD→LSRB→LSRC。

现要求当主 CR-LSP 发生故障时，流量切换到热备份 CR-LSP；当主 CR-LSP 故障恢复，延时 15s 后流量回切到主 CR-LSP。如果主、备 CR-LSP 都发生故障，触发建立逃生路径，使流量可切换到逃生路径上。

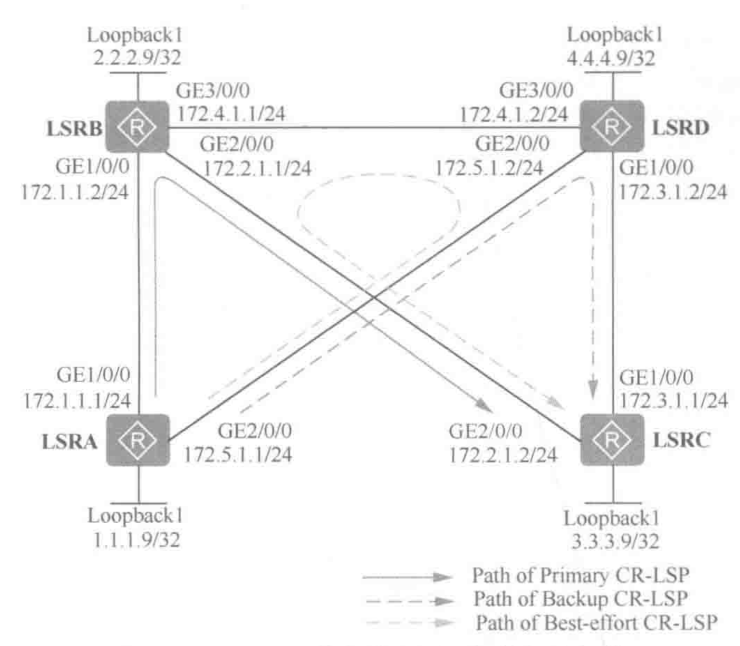


图 7-2　CR-LSP 热备份配置示例的拓扑结构

1. 基本配置思路分析

根据本示例的要求、第 5 章介绍的基本 MPLS TE 隧道配置方法，以及本节前面介绍的备份 CR-LSP 配置方法可得出本示例的如下基本配置思路。

（1）在各节点上配置各接口（包括 Loopback 接口）的 IP 地址和 OSPF 协议，实现各节点之间公网路由可达。

（2）在各节点上配置 LSR ID，并使能各节点全局和公网接口的 MPLS、MPLS TE 和 RSVP-TE 能力，在入节点 LSRA 上使能 CSPF 功能。本示例仅介绍从 LSRA 到 LSRC 的单向 TE 隧道的配置，故入节点仅为 LSRA，实际应用中需要配置双向 TE 隧道，也就是要进行双向 TE 隧道的对应配置。

（3）在各节点上使能 OSPF TE，使得 OSPF 协议可以发布 MPLS TE 信息。

（4）在入节点 LSRA 上配置主、备 CR-LSP 的显式路径。

（5）在入节点 LSRA 上创建目的地址为 LSRC 的隧道接口，指定前面配置的主备 CR-LSP 的显式路径，并使能热备份和逃生路径，配置回切时间为 15s。

2. 具体配置步骤

（1）配置各节点的各接口的 IP 地址和 OSPF 协议。

\#　LSRA 上的配置。

```
<Huawei> system-view
[Huawei] sysname LSRA
[LSRA] interface gigabitethernet 1/0/0
[LSRA-GigabitEthernet1/0/0] ip address 172.1.1.1 255.255.255.0
[LSRA-GigabitEthernet1/0/0] quit
[LSRA] interface gigabitethernet 2/0/0
[LSRA-GigabitEthernet2/0/0] ip address 172.5.1.1 255.255.255.0
[LSRA-GigabitEthernet2/0/0] quit
[LSRA] interface loopback 1
[LSRA-LoopBack1] ip address 1.1.1.9 255.255.255.255
[LSRA-LoopBack1] quit
```

```
[LSRA] ospf 1
[LSRA-ospf-1] area 0
[LSRA-ospf-1-area-0.0.0.0] network 1.1.1.9 0.0.0.0
[LSRA-ospf-1-area-0.0.0.0] network 172.1.1.0 0.0.0.255
[LSRA-ospf-1-area-0.0.0.0] network 172.5.1.0 0.0.0.255
[LSRA-ospf-1-area-0.0.0.0] quit
[LSRA-ospf-1] quit
```

LSRB 上的配置。

```
<Huawei> system-view
[Huawei] sysname LSRB
[LSRB] interface gigabitethernet 1/0/0
[LSRB-GigabitEthernet1/0/0] ip address 172.1.1.2 255.255.255.0
[LSRB-GigabitEthernet1/0/0] quit
[LSRB] interface gigabitethernet 2/0/0
[LSRB-GigabitEthernet2/0/0] ip address 172.2.1.1 255.255.255.0
[LSRB-GigabitEthernet2/0/0] quit
[LSRB] interface gigabitethernet 3/0/0
[LSRB-GigabitEthernet3/0/0] ip address 172.4.1.1 255.255.255.0
[LSRB-GigabitEthernet3/0/0] quit
[LSRB] interface loopback 1
[LSRB-LoopBack1] ip address 2.2.2.9 255.255.255.255
[LSRB-LoopBack1] quit
[LSRB] ospf 1
[LSRB-ospf-1] area 0
[LSRB-ospf-1-area-0.0.0.0] network 2.2.2.9 0.0.0.0
[LSRB-ospf-1-area-0.0.0.0] network 172.1.1.0 0.0.0.255
[LSRB-ospf-1-area-0.0.0.0] network 172.2.1.0 0.0.0.255
[LSRB-ospf-1-area-0.0.0.0] network 172.4.1.0 0.0.0.255
[LSRB-ospf-1-area-0.0.0.0] quit
[LSRB-ospf-1] quit
```

LSRC 上的配置。

```
<Huawei> system-view
[Huawei] sysname LSRC
[LSRC] interface gigabitethernet 1/0/0
[LSRC-GigabitEthernet1/0/0] ip address 172.3.1.1 255.255.255.0
[LSRC-GigabitEthernet1/0/0] quit
[LSRC] interface gigabitethernet 2/0/0
[LSRC-GigabitEthernet2/0/0] ip address 172.2.1.2 255.255.255.0
[LSRC-GigabitEthernet2/0/0] quit
[LSRC] interface loopback 1
[LSRC-LoopBack1] ip address 3.3.3.9 255.255.255.255
[LSRC-LoopBack1] quit
[LSRC] ospf 1
[LSRC-ospf-1] area 0
[LSRC-ospf-1-area-0.0.0.0] network 3.3.3.9 0.0.0.0
[LSRC-ospf-1-area-0.0.0.0] network 172.3.1.0 0.0.0.255
[LSRC-ospf-1-area-0.0.0.0] network 172.2.1.0 0.0.0.255
[LSRC-ospf-1-area-0.0.0.0] quit
[LSRC-ospf-1] quit
```

LSRD 上的配置。

```
<Huawei> system-view
[Huawei] sysname LSRD
[LSRD] interface gigabitethernet 1/0/0
[LSRD-GigabitEthernet1/0/0] ip address 172.3.1.2 255.255.255.0
```

```
[LSRD-GigabitEthernet1/0/0] quit
[LSRD] interface gigabitethernet 2/0/0
[LSRD-GigabitEthernet2/0/0] ip address 172.5.1.2 255.255.255.0
[LSRD-GigabitEthernet2/0/0] quit
[LSRD] interface gigabitethernet 3/0/0
[LSRD-GigabitEthernet3/0/0] ip address 172.4.1.2 255.255.255.0
[LSRD-GigabitEthernet3/0/0] quit
[LSRD] interface loopback 1
[LSRD-LoopBack1] ip address 4.4.4.9 255.255.255.255
[LSRD-LoopBack1] quit
[LSRD] ospf 1
[LSRD-ospf-1] area 0
[LSRD-ospf-1-area-0.0.0.0] network 4.4.4.9 0.0.0.0
[LSRD-ospf-1-area-0.0.0.0] network 172.3.1.0 0.0.0.255
[LSRD-ospf-1-area-0.0.0.0] network 172.4.1.0 0.0.0.255
[LSRD-ospf-1-area-0.0.0.0] network 172.5.1.0 0.0.0.255
[LSRD-ospf-1-area-0.0.0.0] quit
[LSRD-ospf-1] quit
```

以上配置完成后，在各节点上执行 **display ip routing-table** 命令，应可以看到相互之间都学到了到对方 Loopback1 的路由。

（2）配置各节点的 MPLS LSR ID，并使能全局和各公网接口的 MPLS、MPLS TE 和 RSVP-TE 能力，在 LSRA 上使能 CSPF。

\#　LSRA 上的配置。

```
[LSRA] mpls lsr-id 1.1.1.9
[LSRA] mpls
[LSRA-mpls] mpls te
[LSRA-mpls] mpls rsvp-te
[LSRA-mpls] mpls te cspf
[LSRA-mpls] quit
[LSRA] interface gigabitethernet 1/0/0
[LSRA-GigabitEthernet1/0/0] mpls
[LSRA-GigabitEthernet1/0/0] mpls te
[LSRA-GigabitEthernet1/0/0] mpls rsvp-te
[LSRA-GigabitEthernet1/0/0] quit
[LSRA] interface gigabitethernet 2/0/0
[LSRA-GigabitEthernet2/0/0] mpls
[LSRA-GigabitEthernet2/0/0] mpls te
[LSRA-GigabitEthernet2/0/0] mpls rsvp-te
[LSRA-GigabitEthernet2/0/0] quit
```

\#　LSRB 上的配置。

```
[LSRB] mpls lsr-id 2.2.2.9
[LSRB] mpls
[LSRB-mpls] mpls te
[LSRB-mpls] mpls rsvp-te
[LSRB-mpls] quit
[LSRB] interface gigabitethernet 1/0/0
[LSRB-GigabitEthernet1/0/0] mpls
[LSRB-GigabitEthernet1/0/0] mpls te
[LSRB-GigabitEthernet1/0/0] mpls rsvp-te
[LSRB-GigabitEthernet1/0/0] quit
[LSRB] interface gigabitethernet 2/0/0
[LSRB-GigabitEthernet2/0/0] mpls
[LSRB-GigabitEthernet2/0/0] mpls te
```

```
[LSRB-GigabitEthernet2/0/0] mpls rsvp-te
[LSRB-GigabitEthernet2/0/0] quit
[LSRB] interface gigabitethernet 3/0/0
[LSRB-GigabitEthernet3/0/0] mpls
[LSRB-GigabitEthernet3/0/0] mpls te
[LSRB-GigabitEthernet3/0/0] mpls rsvp-te
[LSRB-GigabitEthernet3/0/0] quit
```

LSRC 上的配置。

```
[LSRC] mpls lsr-id 3.3.3.9
[LSRC] mpls
[LSRC-mpls] mpls te
[LSRC-mpls] mpls rsvp-te
[LSRC-mpls] quit
[LSRC] interface gigabitethernet 1/0/0
[LSRC-GigabitEthernet1/0/0] mpls
[LSRC-GigabitEthernet1/0/0] mpls te
[LSRC-GigabitEthernet1/0/0] mpls rsvp-te
[LSRC-GigabitEthernet1/0/0] quit
[LSRC] interface gigabitethernet 2/0/0
[LSRC-GigabitEthernet2/0/0] mpls
[LSRC-GigabitEthernet2/0/0] mpls te
[LSRC-GigabitEthernet2/0/0] mpls rsvp-te
[LSRC-GigabitEthernet2/0/0] quit
```

LSRD 上的配置。

```
[LSRD] mpls lsr-id 4.4.4.9
[LSRD] mpls
[LSRD-mpls] mpls te
[LSRD-mpls] mpls rsvp-te
[LSRD-mpls] quit
[LSRD] interface gigabitethernet 1/0/0
[LSRD-GigabitEthernet1/0/0] mpls
[LSRD-GigabitEthernet1/0/0] mpls te
[LSRD-GigabitEthernet1/0/0] mpls rsvp-te
[LSRD-GigabitEthernet1/0/0] quit
[LSRD] interface gigabitethernet 2/0/0
[LSRD-GigabitEthernet2/0/0] mpls
[LSRD-GigabitEthernet2/0/0] mpls te
[LSRD-GigabitEthernet2/0/0] mpls rsvp-te
[LSRD-GigabitEthernet2/0/0] quit
[LSRD] interface gigabitethernet 3/0/0
[LSRD-GigabitEthernet3/0/0] mpls
[LSRD-GigabitEthernet3/0/0] mpls te
[LSRD-GigabitEthernet3/0/0] mpls rsvp-te
[LSRD-GigabitEthernet3/0/0] quit
```

（3）配置各节点的 OSPF TE 能力，使合节点可通过 OSPF TE 发布 TE 信息。

LSRA 上的配置。

```
[LSRA] ospf
[LSRA-ospf-1] opaque-capability enable
[LSRA-ospf-1] area 0
[LSRA-ospf-1-area-0.0.0.0] mpls-te enable
[LSRA-ospf-1-area-0.0.0.0] quit
[LSRA-ospf-1] quit
```

LSRB 上的配置。

```
[LSRB] ospf
[LSRB-ospf-1] opaque-capability enable
[LSRB-ospf-1] area 0
[LSRB-ospf-1-area-0.0.0.0] mpls-te enable
[LSRB-ospf-1-area-0.0.0.0] quit
[LSRB-ospf-1] quit
```

LSRC 上的配置。

```
[LSRC] ospf
[LSRC-ospf-1] opaque-capability enable
[LSRC-ospf-1] area 0
[LSRC-ospf-1-area-0.0.0.0] mpls-te enable
[LSRC-ospf-1-area-0.0.0.0] quit
[LSRC-ospf-1] quit
```

LSRD 上的配置。

```
[LSRD] ospf
[LSRD-ospf-1] opaque-capability enable
[LSRD-ospf-1] area 0
[LSRD-ospf-1-area-0.0.0.0] mpls-te enable
[LSRD-ospf-1-area-0.0.0.0] quit
[LSRD-ospf-1] quit
```

（4）在入节点 LSRA 上根据要求配置主、备 CR-LSP 使用的显式路径。

在 LSRA 上配置主 CR-LSP 使用的显式路径。

```
[LSRA] explicit-path pri-path
[LSRA-explicit-path-pri-path] next hop 172.1.1.2
[LSRA-explicit-path-pri-path] next hop 172.2.1.2
[LSRA-explicit-path-pri-path] next hop 3.3.3.9
[LSRA-explicit-path-pri-path] quit
```

在 LSRA 上配置备份 CR-LSP 使用的显式路径。

```
[LSRA] explicit-path backup-path
[LSRA-explicit-path-backup-path] next hop 172.5.1.2
[LSRA-explicit-path-backup-path] next hop 172.3.1.1
[LSRA-explicit-path-backup-path] next hop 3.3.3.9
[LSRA-explicit-path-backup-path] quit
```

完成以上配置后，可在 LSRA 上执行 **display explicit-path** 命令查看已经配置的主、备 CR-LSP 显式路径，验证显式路径配置是否正确。

```
[LSRA] display explicit-path pri-path
Path Name : pri-path     Path Status : Enabled
 1        172.1.1.2          Strict        Include
 2        172.2.1.2          Strict        Include
 3        3.3.3.9            Strict        Include

[LSRA] display explicit-path backup-path
Path Name : backup-path    Path Status : Enabled
 1        172.5.1.2          Strict        Include
 2        172.3.1.1          Strict        Include
 3        3.3.3.9            Strict        Include
```

（5）在入节点 LSRA 创建隧道接口，引用前面配置主、备 CR-LSP 的显式路径（上一步已配置好），配置逃生路径和回切时间（即当主 CR-LSP 故障恢复后把流量从备份 CR-LSP 切换到主 CR-LSP 所需等待的时间）。

在 LSRA 上创建 Tunnel 接口，指定主 CR-LSP 显式路径。

```
[LSRA] interface tunnel 0/0/1
[LSRA-Tunnel0/0/1] ip address unnumbered interface loopback 1
[LSRA-Tunnel0/0/1] tunnel-protocol mpls te
[LSRA-Tunnel0/0/1] destination 3.3.3.9
[LSRA-Tunnel0/0/1] mpls te tunnel-id 100
[LSRA-Tunnel0/0/1] mpls te path explicit-path pri-path    #---指定主 CR-LSP 所使用的显式路径的名称为 pri-path
```

在以上 Tunnel 接口配置 CR-LSP 热备份，回切时间为 15s，指定备份 CR-LSP 的显式路径，并配置逃生路径。

```
[LSRA-Tunnel0/0/1] mpls te backup hot-standby wtr 15   #---配置备份 CR-LSP 回切到主 CR-LSP 的延时时间为 15 秒
[LSRA-Tunnel0/0/1] mpls te path explicit-path backup-path secondary   #---指定备份 CR-LSP 使用的显式路径的名称
为 backup-path
[LSRA-Tunnel0/0/1] mpls te backup ordinary best-effort   #--- 配置逃生路径
[LSRA-Tunnel0/0/1] mpls te commit
[LSRA-Tunnel0/0/1] quit
```

以上配置完成后，在 LSRA 上执行 **display mpls te tunnel-interface tunnel 0/0/1** 命令，可发现主 CR-LSP、备份 CR-LSP 建立成功（状态为 Up），但当前活跃 CR-LSP 为主 CR-LSP（Primary LSP），参见输出信息中的粗体字部分。

```
[LSRA] display mpls te tunnel-interface tunnel 0/0/1
    ----------------------------------------------------------------
                        Tunnel0/0/1
    ----------------------------------------------------------------
    Tunnel State Desc          :  Up
    Active LSP                 :  Primary LSP
    Session ID                 :  100
    Ingress LSR ID             :  1.1.1.9          Egress LSR ID    : 3.3.3.9
    Admin State                :  Up                 Oper State     : Up
    Primary LSP State          : Up
      Main LSP State           : READY               LSP ID         : 10
    Hot-Standby LSP State      : Up
      Main LSP State           : READY               LSP ID         : 32773
```

在 LSRA 上使用 **display mpls te hot-standby state interface Tunnel 0/0/1** 命令查看热备份信息。

```
[LSRA] display mpls te hot-standby state interface Tunnel 0/0/1
----------------------------------------------------------------
Verbose information about the Tunnel0/0/1 hot-standby state
----------------------------------------------------------------
session id                          : 100
main LSP token                      : 0xc
hot-standby LSP token               : 0xb
HSB switch result                   : Primary LSP
HSB switch reason                   : -
WTR config time                     : 15s
WTR remain time                     : -
using overlapped path               : no
```

可在 LSRA 上使用 **ping lsp te** 命令检测热备份 CR-LSP 的连通性。

```
[LSRA] ping lsp te tunnel 0/0/1 hot-standby
  LSP ping FEC: TE TUNNEL IPV4 SESSION QUERY Tunnel0/0/1 : 100   data bytes, pres
s CTRL_C to break
    Reply from 3.3.3.9: bytes=100 Sequence=1 time=11 ms
    Reply from 3.3.3.9: bytes=100 Sequence=2 time=2 ms
    Reply from 3.3.3.9: bytes=100 Sequence=3 time=2 ms
    Reply from 3.3.3.9: bytes=100 Sequence=4 time=2 ms
```

```
    Reply from 3.3.3.9: bytes=100 Sequence=5 time=2 ms

  --- FEC: TE TUNNEL IPV4 SESSION QUERY Tunnel0/0/1 ping statistics ---
    5 packet(s) transmitted
    5 packet(s) received
    0.00% packet loss
    round-trip min/avg/max = 2/3/11 ms
```

或者在 LSRA 上使用 **tracert lsp te** 命令检测热备份 CR-LSP 所经过的路径。

```
[LSRA] tracert lsp te tunnel 0/0/1 hot-standby
  LSP Trace Route FEC: TE TUNNEL IPV4 SESSION QUERY Tunnel0/0/1 , press CTRL_C t
o break.
  TTL    Replier          Time     Type      Downstream
  0                                Ingress   172.5.1.2/[1027 ]
  1      172.5.1.2        9 ms     Transit   172.3.1.1/[3 ]
  2      3.3.3.9          10 ms    Egress
```

3. 配置结果验证

在 LSRA 的 GE1/0/0 接口上执行 **shutdown** 命令，模拟线缆拔出，主 CR-LSP 出现故障。然后在 LSRA 上执行 **display mpls te tunnel-interface tunnel** 0/0/1 命令，会发现流量被切换到备份 CR-LSP 上，因为显示当前活跃的 LSP 为备份 CR-LSP 了，参见输出信息中的粗体字部分。

```
[LSRA] display mpls te tunnel-interface tunnel 0/0/1
    ----------------------------------------------------------------
                         Tunnel0/0/1
    ----------------------------------------------------------------
    Tunnel State Desc    :  Up
    Active LSP           :  Hot-Standby LSP
    Session ID           :  100
    Ingress LSR ID       :  1.1.1.9          Egress LSR ID:  3.3.3.9
    Admin State          :  Up               Oper State   :  Up
    Primary LSP State      : DOWN
      Main LSP State       : SETTING Up
    Hot-Standby LSP State  : Up
      Main LSP State       : READY                LSP ID   : 32773
```

在 LSRA 的 GE1/0/0 接口执行 **undo shutdown** 命令后，等 15s（配置的回切时间）后，可发现流量又被切换到主 CR-LSP 上。

如果关闭 LSRA 或 LSRB 上的 GE1/0/0 接口后，再关闭 LSRC 或 LSRD 上的 GE1/0/0 接口（同时关闭了备份 CR-LSP），隧道接口会先变为 Down，随后又会变为 Up，因为逃生路径建立成功，流量被切换到逃生路径上。此时在 LSRA 上执行 **display mpls te tunnel-interface tunnel 0/0/1** 命令，发现隧道接口变 Down，逃生路径正在建立。

```
[LSRA] display mpls te tunnel-interface tunnel 0/0/1
    ----------------------------------------------------------------
                         Tunnel0/0/1
    ----------------------------------------------------------------
    Tunnel State Desc    :  DOWN
    Active LSP           :  -
    Session ID           :  100
    Ingress LSR ID       :  1.1.1.9          Egress LSR ID:  3.3.3.9
    Admin State          :  Up               Oper State   :  DOWN
    Primary LSP State      : DOWN
      Main LSP State       : SETTING Up
```

```
Hot-Standby LSP State   : DOWN
  Main LSP State          : SETTING Up
Best-Effort LSP State   : DOWN
  Main LSP State          : SETTING Up
```

几秒之后，在 LSRA 上再次执行 **display mpls te tunnel-interface tunnel 0/0/1** 命令，发现隧道接口 Up，逃生路径建立成功，参见输出信息中的粗体字部分。

```
[LSRA] display mpls te tunnel-interface tunnel 0/0/1
----------------------------------------------------------------
                        Tunnel0/0/1
----------------------------------------------------------------
Tunnel State Desc   :  Up
Active LSP          :  Best-Effort LSP
Session ID          :  100
Ingress LSR ID      :  1.1.1.9          Egress LSR ID:  3.3.3.9
Admin State         :  Up               Oper State   :  Up
Primary LSP State       : DOWN
  Main LSP State          : SETTING Up
Hot-Standby LSP State   : DOWN
  Main LSP State          : SETTING Up
Best-Effort LSP State   : Up
  Main LSP State          : READY               LSP ID   : 32776
```

7.2 BFD for MPLS TE 配置与管理

MPLS TE 经常采用 TE FRR、CR-LSP 备份或 TE 隧道保护组来提高网络的可靠性，但是这几种技术依靠 RSVP Hello 或者 RSVP 消息刷新超时等机制进行故障检测，检测速度缓慢。当节点间存在二层设备（比如二层交换机）时，触发流量保护倒换的速度将变慢，一定程度上引起流量的丢失。BFD 检测机制可以很好地解决这个问题，它采用快速收发报文的机制，完成这些隧道链路故障的快速检测，从而引导承载的业务流量进行快速切换，达到保护业务的目的。

7.2.1 BFD for MPLS TE 简介

MPLS TE 中的 BFD 检测技术按照检测对象不同可分为三种：BFD for RSVP、BFD for CR-LSP 和 BFD for TE Tunnel。

1. BFD for RSVP

BFD for RSVP 是对 RSVP 的检测，可实现毫秒级故障监测时间，配合 RSVP 协议可快速地发现 RSVP 邻接故障。BFD for RSVP 一般用在 TE FRR 中 PLR（本地修复节点）节点与主路径的 RSVP 邻居之间存在二层设备的情况，如图 7-3 所示（中间节点为二层交换机）。

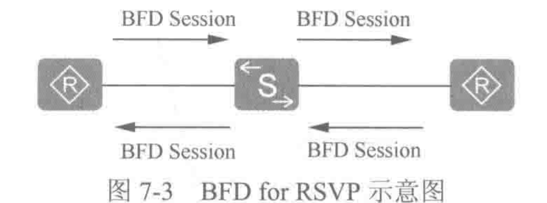


图 7-3 BFD for RSVP 示意图

BFD for RSVP 可以与 BFD for OSPF、BFD for ISIS 和 BFD for BGP 共享会话。共享 BFD 会话时，则本地节点分别选择所有共享 BFD 会话的协议发送时间间隔、接收时间间隔、本地检测倍数的最小值作为本地的 BFD 会话参数。

2. BFD for CR-LSP

BFD for CR-LSP 是对 CR-LSP 的检测，能够快速检测到 CR-LSP 的故障，并及时通知转发平面，从而保证流量的快速切换。BFD for CR-LSP 通常与热备份 CR-LSP 配合使用。如图 7-4 所示，配置好 BFD for CR-LSP 后，在入节点和出节点之间就会建立 BFD 会话。BFD 报文从源端开始经过 CR-LSP 转发到达目的端，目的端再对该 BFD 报文进行回应，通过此方式在源端可以快速检测出 CR-LSP 所经过链路的状态。

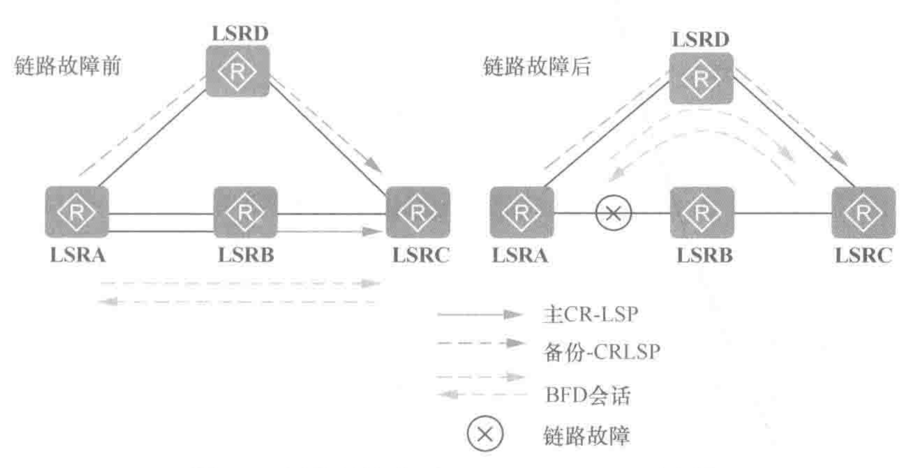


图 7-4 链路故障前后 BFD for CR-LSP 示意图

当检测出链路故障以后，BFD 将此信息上报给设备转发平面。转发平面查找备份 CR-LSP，将业务流量切换到备份 CR-LSP 上，然后设备转发平面再将故障信息上报给控制平面。如果采用的是动态 BFD for CR-LSP，控制平面会主动去创建备份 CR-LSP 的 BFD 会话；如果采用的是静态 BFD for CR-LSP 时，且需要对备份 CR-LSP 进行检测，则可以为其配置 BFD 检测。

3. BFD for TE Tunnel

BFD for TE Tunnel 用于 MPLS TE 隧道作为 VPN 的公网隧道时的场景，可使用 BFD 检测整条 TE 隧道，从而触发 VPN FRR 等应用进行流量切换。

BFD for TE Tunnel 与 BFD for CR-LSP 的区别是故障通告的对象及故障后切换的对象不同：BFD for TE Tunnel 是向 VPN 等应用通告故障，触发业务流在不同 TE 隧道的切换；BFD for CR-LSP 是向 TE 隧道通告故障，触发业务流在同一 TE 隧道内的不同 CR-LSP 上的切换。

BFD for RSVP、BFD for CR-LSP 和 BFD for TE Tunnel 的区别见表 7-5。由于篇幅原因，本节后面仅介绍最常用的 BFD for CR-LSP 的配置方法。

表 7-5 **BFD for TE 中三种不同检测技术的比较**

检测技术	检测对象	部署位置	适用场景	BFD 会话方式支持
BFD for RSVP	RSVP 邻居关系	RSVP 会话的两个邻居节点	与 TE FRR 联用	动态

（续表）

检测技术	检测对象	部署位置	适用场景	BFD 会话方式支持
BFD for CR-LSP	CR-LSP	隧道的入/出节点	与热备份 CR-LSP 联用	动态和静态
BFD for TE Tunnel	MPLS TE 隧道	隧道的入/出节点	与 VPN FRR 或者 VLL FRR 联用	静态

7.2.2 静态 BFD for CR-LSP 配置与管理

所谓“静态 BFD for CR-LSP”就是采用静态配置的 BFD 会话来检测 CR-LSP（可以是主 CR-LSP，也可以是备份 CR-LSP）的连通性。主要是利用 BFD 检测的快速特性，当检测到主 CR-LSP 发生故障时，可及时通知转发平面快速把流量转发路径切换到备份 CR-LSP 上。

静态 BFD for CR-LSP 的配置就是需要配置静态 BFD 会话中 TE 隧道入/ 出节点相关的参数，当然首先是要在 TE 隧道入/ 出节点上全局使能 BFD 功能。总体来说就是可分两部分：（1）配置入节点 BFD 参数，（2）配置出节点 BFD 参数。

以上两项配置任务的基本配置方法与我们在第 2 章介绍的静态 BFD 检测静态 LSP 的配置方法基本一样，只是这里所绑定的对象不是普通 MPLS 隧道中的 LSP，而是 MPLS TE 隧道中的 CR-LSP。

配置静态 BFD for CR-LSP 前，需要完成所应用的静态/动态 MPLS TE 隧道、静态/动态 MPLS DS-TE 隧道，或 CR-LSP 备份的配置。

1．配置入节点的 BFD 参数

可以在入节点上按照表 7-6 所示步骤配置静态 BFD 会话参数，包括本地标识符、远端标识符、本地发送 BFD 报文的时间间隔、本地允许接收 BFD 报文的时间间隔和本地 BFD 检测倍数等，这些将会影响会话的建立。

表 7-6 配置入节点的 **BFD** 参数的步骤

步骤	命令	说明
1	**system-view** 例如：<Huawei> **system-view**	进入系统视图
2	**bfd** 例如：[Huawei] **bfd**	使能全局 BFD 能力并进入 BFD 全局视图。 缺省情况下，全局 BFD 功能未使能，可用 **undo bfd** 命令全局去使能 BFD 功能，此时如果已经配置了 BFD 会话信息，则所有的 BFD 会话都会被删除
3	**quit** 例如：[Huawei-bfd] **quit**	返回系统视图
4	**bfd** *cfg-name* **bind mpls-te interface tunnel** *interface-number* **te-lsp** [**backup**] 例如：[Huawei] **bfd** 1to4 **bind mpls-te interface Tunnel** 0/0/1 **te-lsp backup**	配置 BFD 会话绑定指定 Tunnel 的主用或备用 CR-LSP。命令中的参数说明如下。 • *cfg-name*：指定 BFD 配置名，字符串形式，不支持空格，不区分大小写，长度范围是 1～15。当输入的字符串两端使用双引号时，可在字符串中输入空格。 • **interface tunnel** *interface-number*：指定 BFD 会话绑定的 Tunnel 接口编号。 • **te-lsp**：指定 BFD 检测的对象是所绑定 Tunnel 隧道中的 CR-LSP，如果没有这个关键字，则表示 BFD 的是 TE 隧道。

（续表）

步骤	命令	说明
4	**bfd** *cfg-name* **bind mpls-te interface tunnel** *interface-number* **te-lsp** [**backup**] 例如：[Huawei] **bfd** 1to4 **bind mpls-te interface Tunnel** 0/0/1 **te-lsp backup**	• **backup**：可选项，指定 BFD 检测的是备份 CR-LSP。如果不选择本可选项，则 BFD 检测的是主 CR-LSP。 缺省情况下，Tunnel 隧道没有使用 BFD 检测，可用 **undo bfd** *cfg-name* 命令删除指定的 BFD 会话
5	**discriminator local** *discr-value* 例如：[Huawei-bfd-session-1to4] **discriminator local** 10	配置本地标识符，整数形式，取值范围是 1～8191。 【注意】BFD 会话两端设备的本地标识符和远端标识符需要分别对应，即本端的本地标识符与对端的远端标识符相同，否则会话无法正确建立。并且，本地标识符和远端标识符配置成功后不可修改，如果需要修改静态 BFD 会话本地标识符或者远端标识符，则必须先删除该 BFD 会话，然后再配置本地标识符
6	**discriminator remote** *discr-value* 例如：[Huawei-bfd-session-1to4] **discriminator remote** 20	配置远端标识符，整数形式，取值范围是 1～8191。其他说明参见上面第 5 步
7	**min-tx-interval** *interval* 例如：[Huawei-bfd-session-1to4] **min-tx-interval** 300	（可选）调整本地发送 BFD 报文的时间间隔，整数形式，取值范围是 10～2000，单位是毫秒。 如果 BFD 会话在设置的检测周期内没有收到对端发来的 BFD 报文，则认为链路发生了故障，BFD 会话的状态将会置为 Down。为降低对系统资源的占用，一旦检测到 BFD 会话状态变为 Down，系统自动将本端的发送间隔调整为大于 1000ms 的随机值，当 BFD 会话的状态重新变为 Up 后，再恢复成用户配置的时间间隔。 【说明】用户可以根据网络的实际状况增大或者降低 BFD 报文的发送和接收时间间隔。BFD 报文的发送、接收时间间隔直接决定了 BFD 会话的检测时间。对于不太稳定的链路，如果配置的 BFD 报文的发送、接收时间间隔较小，则 BFD 会话可能会发生震荡，这时可以选择增大 BFD 报文的发送和接收时间间隔。通常情况下，建议使用缺省值。 缺省情况下，发送间隔是 1000ms，可用 **undo min-tx-interval** 命令恢复 BFD 报文的发送间隔为缺省值
8	**min-rx-interval** *interval* 例如：[Huawei-bfd-session-1to4] **min-rx-interval** 600	（可选）调整本地接收 BFD 报文的时间间隔，整数形式，取值范围是 10～2000，单位是毫秒。其他说明参见上面第 7 步。 缺省情况下，接收间隔是 1000ms，可用 **undo min-rx-interval** 命令恢复 BFD 报文的接收间隔为缺省值
9	**detect-multiplier** *multiplier* 例如：[Huawei-bfd-session-1to4] **detect-multiplier** 5	（可选）调整本地 BFD 检测倍数，整数形式，取值范围是 3～50。 BFD 会话的本端检测倍数直接决定了对端 BFD 会话的检测时间，检测时间 = 接收到的远端 Detect Multi（检测倍数）× max（本地的 RMRI，接收到的 DMTI）。其中，Detect Mult 是检测倍数，通过本条命令配置；RMRI 是本端能够支持的最短 BFD 报文接收间隔，是通过第 8 步 **min-rx-**

（续表）

步骤	命令	说明
9	**detect-multiplier** *multiplier* 例如：[Huawei-bfd-session-1to4] **detect-multiplier** 5	**interval** *interval* 命令配置的；DMTI 是本端想要采用的最短 BFD 报文的发送间隔，是通过第 7 步 **min-tx-interval** *interval* 命令配置的。 缺省情况下，本地 BFD 检测倍数为 3，可用 **undo detect-multiplier** 命令恢复 BFD 会话的本地检测倍数为缺省值
10	**process-pst** 例如：[Huawei-bfd-session-1to4] **process-pst**	使能系统在 BFD 会话状态变化时修改端口状态表功能。该命令的功能是 BFD 会话状态变化时通知应用协议进行主/备 CR-LSP 之间的快速切换。 缺省情况下，修改端口状态表 PST 功能处于未使能状态，可用 **undo process-pst** 命令恢复缺省配置
11	**notify neighbor-down** 例如：[Huawei-bfd-session-1to4] **notify neighbor-down**	设置 BFD 会话检测到邻居 Down 故障时通知上层协议。出现以下任何一种情况均会通知上层协议。 • BFD 会话在检测时间超时后通知上层协议：BFD 会话需要在两端配置，如果一端的 BFD 会话没有收到对端发来的 BFD 报文，则会认为链路发生了故障，此时 BFD 会话将此故障信息通知给上层协议，延时比较长。 • BFD 会话在检测到邻居 Down 后通知上层协议：配置 BFD 会话一端检测到了邻居 Down，则此时无需等到检测超时，而是直接将邻居 Down 的故障信息通知上层协议，延时比较短。 对于 BFD 检测 CR-LSP 链路，用户只关心本端到对端链路是否故障，无需关注对端到本端的链路状态。所以只需通过本命令设置 BFD 会话检测到的邻居 Down 故障通知上层协议，从而防止由于采用“BFD 会话在检测时间超时通知上层协议”功能才进行业务切换，影响正常业务。 缺省情况下，BFD 会话在检测时间超时或者检测到邻居 Down 后均通知上层协议，可用 **undo notify neighbor-down** 命令恢复 BFD 会话检测到故障时通知上层协议的方式为缺省情况
12	**commit** 例如：[Huawei-bfd-session-1to4] **commit**	提交隧道配置，使配置更改生效

2. 配置出节点 BFD 参数

可以在出节点上按照表 7-7 所示步骤配置的 BFD 参数，也包括本地标识符、远端标识符、本地发送 BFD 报文的时间间隔、本地允许接收 BFD 报文的时间间隔和本地 BFD 检测倍数等，这些将会影响会话的建立。但这里与入节点的配置有一个最重要的区别就是出节点向入节点通告故障的反向通道可以有多种选择，不一定也是 CR-LSP，要根据具体情形来选择。但为了保证 BFD 报文往返路径一致，一般情况下反向通道优先选用 CR-LSP。

表 7-7 配置出节点 BFD 参数的步骤

步骤	命令	说明
1	**system-view** 例如：<Huawei> **system-view**	进入系统视图

（续表）

步骤	命令	说明
2	**bfd** 例如：[Huawei] **bfd**	使能全局 BFD 能力并进入 BFD 全局视图。其他说明参见表 7-22 中的第 2 步
3	**quit** 例如：[Huawei-bfd] **quit**	返回系统视图
4	**bfd** *cfg-name* **bind peer-ip** *peer-ip* [**vpn-instance** *vpn-instance-name*] [**interface** *interface-type interface-number*] [**source-ip** *source-ip*] 例如：[Huawei] **bfd** 1to4 **bind peer-ip** 10.10.20.2	（四选一）当反向通道是 IP 链路时创建 BFD 会话。在创建 BFD 会话时，单跳检测必须绑定对端 IP 地址和本端相应接口，多跳检测只需绑定对端 IP 地址。命令中的参数说明如下。 • *cfg-name*：指定 BFD 配置名，字符串形式，不支持空格，不区分大小写，长度范围是 1～15。当输入的字符串两端使用双引号时，可在字符串中输入空格。 • **peer-ip** *ip-address*：指定 BFD 会话绑定的对端 IP 地址。**如果只指定对端 IP 地址，则表示检测多跳链路**。 • **vpn-instance** *vpn-name*：可选参数，指定对端 BFD 会话绑定的 VPN 实例名称，必须是已创建的 VPN 实例。如果不指定 VPN 实例，则认为对端地址是公网地址。如果同时指定了对端 IP 地址和 VPN 实例，则表示检测 VPN 路由的多跳链路。 • **interface** *interface-type interface-number*：可选参数，指定绑定 BFD 会话的接口。**如果同时指定了对端 IP 地址和本端接口，表示检测单跳链路**，即检测以该接口为出接口、以 peer-ip 为下一跳地址的一条固定路由；如果同时指定了对端 IP 地址、VPN 实例和本端接口，表示检测 VPN 路由的单跳链路。 • **source-ip** *ip-address*：可选参数，指定 BFD 报文携带的源 IP 地址。通常情况下，不需要配置该参数。在 BFD 会话协商阶段，如果不配置该参数，则系统将在本地路由表中查找去往对端 IP 地址的出接口，以该出接口的 IP 地址作为本端发送 BFD 报文的源 IP 地址；在 BFD 会话检测链路阶段，如果不配置该参数，则系统会将 BFD 报文的源 IP 地址设置为一个固定的值。 缺省情况下，没有创建 BFD 会话，可用 **undo bfd session-name** 命令删除指定的 BFD 会话，同时取消 BFD 会话的绑定信息
	bfd *cfg-name* **bind static-lsp** *lsp-name* 例如：[Huawei] **bfd** 1to4 **bind static-lsp** 1to4	（四选一）当反向通道是静态 LSP 时创建静态 LSP 的 BFD 会话。命令中的参数说明如下： • *cfg-name*：指定 BFD 配置名。 • *lsp-name*：指定 BFD 会话绑定静态 LSP 的名称，必须是已存在的静态 LSP 名称。 缺省情况下，没有创建检测静态 LSP 的 BFD 会话，可用 **undo bfd** *cfg-name* 命令用来删除指定的 BFD 会话

（续表）

步骤	命令	说明
4	**bfd** *cfg-name* **bind ldp-lsp peer-ip** *ip-address* **nexthop** *ip-address* [**interface** *interface-type interface-number*] 例如：Huawei] **bfd** 1to4 **bind ldp-lsp peer-ip** 4.4.4.4 **nexthop** 1.1.1.1 **interface** gigabitethernet 1/0/0	（四选一）当反向通道是动态 LSP 时创建 LDP LSP 的 BFD 会话。命令中的参数说明如下。 • *cfg-name*：指定 BFD 会话名称。 • **peer-ip** *ip-address*：指定 BFD 会话绑定动态 LDP LSP 的目的端 IP 地址。 • **nexthop** *ip-address*：指定被检测 LSP 的下一跳 IP 地址。 • **interface** *interface-type interface-number* ：可选参数，指定 BFD 绑定的出接口。 缺省情况下，没有创建检测 LDP LSP 的 BFD 会话，可用 **undo bfd** *cfg-name* 命令删除指定的 BFD 会话
	bfd *cfg-name* **bind mpls-te interface tunnel** *interface-number* [**te-lsp** [**backup**]] 例如：[Huawei] **bfd** 1to 4 **bind mpls-te interface** Tunnel 0/0/1 **te-lsp**	（四选一）当反向通道是 CR-LSP 或 TE 隧道时，配置 BFD检测TE隧道或与TE隧道绑定的主用或备用LSP。命令中的参数和选项说明参见表 7-6 中的第 4 步。 BFD 检测 TE 隧道时，如果 TE 隧道的状态为 Down，则能够创建 BFD 会话，但 BFD 会话不能 Up。一个 TE 隧道可能有多个 CR-LSP，当 BFD 检测 TE 隧道时，只有全部 CR-LSP 都出现故障时，BFD 会话的状态才为 Down。 缺省情况下，Tunnel 隧道没有使用 BFD 检测，可用 **undo bfd** *cfg-name* 命令删除指定的 BFD 会话
5	**discriminator local** *discr-value* 例如：[Huawei-bfd-session-1to4] **discriminator local** 10	配置本地标识符，参见表 7-6 中的第 5 步
6	**discriminator remote** *discr-value* 例如：[Huawei-bfd-session-1to4] **discriminator remote** 20	配置远端标识符，参见表 7-6 中的第 6 步
7	**min-tx-interval** *interval* 例如：[Huawei-bfd-session-1to4] **min-tx-interval** 300	（可选）调整本地发送 BFD 报文的时间间隔，参见表 7-6 中的第 7 步
8	**min-rx-interval** *interval* 例如：[Huawei-bfd-session-1to4] **min-rx-interval** 600	（可选）调整本地接收 BFD 报文的时间间隔，参见表 7-6 中的第 8 步
9	**detect-multiplier** *multiplier* 例如：[Huawei-bfd-session-1to4] **detect-multiplier** 5	（可选）调整本地 BFD 检测倍数，参见表 7-6 中的第 9 步
10	**process-pst** 例如：[Huawei-bfd-session-1to4] **process-pst**	（可选）使能系统在 BFD 会话状态变化时修改端口状态表功能，参见表 7-6 中的第 10 步
11	**notify neighbor-down**	设置 BFD 会话检测到邻居 Down 故障时通知上层协议
12	**commit** 例如：[Huawei-bfd-session-1to4] **commit**	提交隧道配置，使配置更改生效

7.2.3　配置动态 BFD for CR-LSP

相对静态 BFD for CR-LSP 而言，配置动态 BFD for CR-LSP 可以减少配置的复杂性，减少人为的配置错误。主要的配置任务如下。

（1）使能入节点动态创建 BFD 会话。

（2）使能出节点被动创建 BFD 会话。

（3）（可选）调整入节点 BFD 检测参数。

因为动态 BFD for CR-LSP 也是对 CR-LSP 的检测，所以在配置前也需要完成所应用的静态/动态 MPLS TE 隧道、静态/动态 MPLS DS-TE 隧道，或 CR-LSP 备份配置。

1. 使能入节点动态创建 BFD 会话

使能 TE 动态创建 BFD 会话有两种方式。

■ 全局使能动态创建 BFD 会话

当入节点的大部分 TE 隧道都需要使能自动创建 BFD 会话时，建议选择该方式。具体配置步骤见表 7-8。

■ Tunnel 接口下使能动态创建 BFD 会话

当入节点的小部分 TE 隧道需要使能自动创建 BFD 会话时，建议选择该方式。具体配置步骤见表 7-9。

表 7-8　全局使能动态创建 **BFD** 会话的配置步骤

步骤	命令	说明
1	**system-view** 例如：<Huawei> **system-view**	进入系统视图
2	**bfd** 例如：[Huawei] **bfd**	使能全局 BFD 能力并进入 BFD 全局视图。其他说明参见表 7-6 中的第 2 步
3	**quit** 例如：[Huawei-bfd] **quit**	返回系统视图
4	**mpls** 例如：[Huawei] **mpls**	进入 MPLS 视图
5	**mpls te bfd enable** 例如：[Huawei-mpls] **mpls te bfd enable**	触发 MPLS TE 自动创建 BFD 会话。配置该命令后，所有 Tunnel 接口都使能了 BFD for TE，除非 Tunnel 接口的 BFD for TE 能力已被阻塞 缺省情况下，未使能 BFD for TE 能力，可用 **undo mpls te bfd enable** 命令恢复缺省配置
6	**quit** 例如：[Huawei-mpls] **quit**	返回系统视图
以下均为可选配置		
7	**interface tunnel** *interface-number* 例如：[Huawei] **interface tunnel** 0/0/1	进入要阻塞 BFD 会话能力的 TE 隧道接口的接口视图
8	**mpls te bfd block** 例如：[Huawei-Tunnel0/0/1] **mpls te bfd block**	阻塞该 TE 隧道自动创建 BFD 会话能力。 缺省情况下，未阻塞 Tunnel 接口的 BFD 能力，可用 **undo mpls te bfd block** 命令恢复为缺省配置
9	**mpls te commit** 例如：[Huawei-Tunnel0/0/1] **mpls te commit**	提交配置，使配置生效

表 7-9　　**Tunnel** 接口下使能动态创建 **BFD** 会话的配置步骤

步骤	命令	说明
1	**system-view** 例如：<Huawei> **system-view**	进入系统视图
2	**bfd** 例如：[Huawei] **bfd**	使能全局 BFD 能力并进入 BFD 全局视图。其他说明参见表 7-6 中的第 2 步
3	**quit** 例如：[Huawei-bfd] **quit**	返回系统视图
4	**interface tunnel** *interface-number* 例如：[Huawei] **interface tunnel** 0/0/1	进入要使能 BFD 会话能力的 TE 隧道接口的接口视图
5	**mpls te bfd enable** 例如：[Huawei-Tunnel0/0/1] **mpls te bfd block**	触发该 TE 隧道自动创建 BFD 会话。 在 Tunnel 接口视图下配置该命令只对当前 Tunnel 接口生效。 缺省情况下，未使能 Tunnel 接口的 BFD 能力，可用 **undo mpls te bfd enable** 命令恢复为缺省配置
6	**mpls te commit** 例如：[Huawei-Tunnel0/0/1] **mpls te commit**	提交配置，使配置生效

2. 使能出节点被动创建 BFD 会话

由于 CR-LSP 路径是单向的，在一条 CR-LSP 路径上，当主动方（源端）创建 BFD 会话后触发 LSP ping 报文发送，被动方（宿端）收到 ping 报文后才可能自动创建 BFD 会话。所以在出节点只需要使能被动创建 BFD 会话功能即可，具体配置步骤见表 7-10。

表 7-10　　使能出节点被动创建 **BFD** 会话的配置步骤

步骤	命令	说明
1	**system-view** 例如：<Huawei> **system-view**	进入系统视图
2	**bfd** 例如：[Huawei] **bfd**	使能全局 BFD 能力并进入 BFD 全局视图。其他说明参见表 7-6 中的第 2 步
3	**mpls-passive** 例如：[Huawei-bfd] **mpls-passive**	使能被动创建 BFD 会话功能。执行完该命令不创建 BFD 会话，而是等接收到源端发送的携带 BFD TLV 的 LSP ping 请求报文后才建立 BFD 会话。 缺省情况下，不使能被动动态创建 BFD 会话功能，可用 **undo mpls-passive** 命令在 LSP 的目的端设备上禁止被动动态创建 BFD 会话的功能

7.2.4 调整入节点 BFD 检测参数

调整隧道入节点的 BFD 检测参数是一项可选项，一般直接采用缺省配置即可。调整的方式有以下两种。

（1）调整全局的 BFD 检测参数

当入节点的大部分 TE 隧道都使用相同的 BFD 检测参数时使用此方式，具体的配置步骤见表 7-11。

（2）调整 Tunnel 接口的 BFD 检测参数

当入节点有些 TE 隧道需要使用与全局不同的 BFD 检测参数时，则在这些隧道的 Tunnel 接口下单独调整 BFD 检测参数，具体的配置步骤见表 7-12。

缺省情况下，本地 BFD 报文实际发送时间间隔 = MAX { 本地配置的发送时间间隔，对端配置的接收时间间隔 }；本地实际接收时间间隔 = MAX { 对端配置的发送时间间隔，本地配置的接收时间间隔 }；本地实际检测时间 = 本地实际接收时间间隔×对端配置的 BFD 检测倍数。**但对于被动创建 BFD 会话的 TE 隧道出节点，BFD 报文的接收时间间隔、发送时间间隔和检测倍数都不能调整，取值是设备可设置的最小值。**因此，BFD for CR-LSP 中，隧道两端节点最终采用的 BFD 检测时间为。

- 入节点实际检测时间间隔 = 入节点配置的接收时间间隔×3。
- 出节点实际检测时间间隔 = 入节点配置的发送时间间隔×入节点配置的检测倍数。

表 7-11　调整全局的 BFD 检测参数的配置步骤

步骤	命令	说明
1	**system-view** 例如：<Huawei> **system-view**	进入系统视图
2	**mpls** 例如：[Huawei] **mpls**	进入 MPLS 视图
3	**mpls te bfd** { **min-tx-interval** *tx-interval* | **min-rx-interval** *rx-interval* | **detect-multiplier** *multiplier* } * 例如：[Huawei-mpls] **mpls te bfd min-tx-interval** 200 **detect-multiplier** 5	设置 BFD 检测的时间参数。命令中的参数说明如下。 • **min-tx-interval** *tx-interval*：可多选参数，指定 BFD 会话发送时间间隔，取值范围是 10～2000，单位是毫秒，缺省值是 1000ms。 • **min-rx-interval** *rx-interval*：可多选参数，指定 BFD 会话接收时间间隔，取值范围是 10～2000，单位是毫秒，缺省值是 1000ms。 • **detect-multiplier** *multiplier*：可多选参数，指定 BFD 会话本地检测的倍数，整数形式，取值范围是 3～50，缺省值是 3。 缺省情况下，没有配置 BFD for TE 会话参数，可用 **undo mpls te bfd** { **min-tx-interval** *tx-interval* | **min-rx-interval** *rx-interval* | **detect-multiplier** *multiplier* } *命令恢复指定参数为缺省配置

表 7-12　调整 Tunnel 接口的 BFD 检测参数的配置步骤

步骤	命令	说明
1	**system-view** 例如：<Huawei> **system-view**	进入系统视图
2	**interface tunnel** *interface-number* 例如：[Huawei] **interface tunnel** 0/0/1	进入 Tunnel 接口视图
3	**mpls te bfd** { **min-tx-interval** *tx-interval* | **min-rx-interval** *rx-interval* | **detect-multiplier** *multiplier* } * 例如：[Huawei-Tunnel0/0/1] **mpls te bfd min-tx-interval** 200 **detect-multiplier** 5	设置 BFD 检测的时间参数。命令中的参数说明参见表 7-11 中的第 3 步。 当入节点配置的 **min-tx-interval** *tx-interval* 和出节点缺省的发送时间间隔 1000ms 不同时，取两者中的较大值为实际的会话参数；实际的 **detect-multiplier** *multiplier* 为出节点上缺省的检测倍数——3

（续表）

步骤	命令	说明
4	**mpls te commit** 例如：[Huawei-Tunnel0/0/1] **mpls te commit**	提交配置，使配置生效

7.2.5 BFD for CR-LSP 配置管理

已经完成静态 BFD for CR-LSP 功能的所有配置后，执行以下 **display** 命令查看相关配置，验证配置效果。

■ **display bfd configuration mpls-te interface tunnel** *interface-number* **te-lsp** [**verbose**]：查看隧道入节点查看 BFD 配置信息。

■ 执行以下命令查看隧道出节点查看 BFD 配置信息。

- **display bfd configuration all** [**for-ip** | **for-lsp** | **for-te**] [**verbose**]：查看所有 BFD 相关配置信息。
- **display bfd configuration static** [**for-ip** | **for-lsp** | **for-te** | **name** *cfg-name*] [**verbose**]：查看静态 BFD 相关配置信息。
- **display bfd configuration peer-ip** *peer-ip* [**vpn-instance** *vpn-instance-name*] [**verbose**]：查看反向通道为 IP 的 BFD 配置信息。
- **display bfd configuration static-lsp** *lsp-name* [**verbose**]：查看反向通道为静态 LSP 的 BFD 配置信息。
- **display bfd configuration ldp-lsp peer-ip** *peer-ip* **nexthop** *nexthop-address* [**interface** *interface-type interface-number*] [**verbose**]：查看反向通道为 LDP LSP 的 BFD 配置信息。
- **display bfd configuration mpls-te interface tunnel** *interface-number* **te-lsp** [**verbose**]：查看反向通道为 CR-LSP 的 BFD 配置信息。
- **display bfd configuration mpls-te interface tunnel** *interface-number* [**verbose**]：查看反向通道为 TE 隧道的 BFD 配置信息。

■ **display bfd session mpls-te interface tunnel** *interface-number* **te-lsp** [**verbose**]：查看隧道入节点查看 BFD 会话信息。

■ 执行以下命令查看隧道出节点查看 BFD 会话信息。

- **display bfd session all** [**for-ip** | **for-lsp** | **for-te**] [**verbose**]：查看所有 BFD 相关配置信息。
- **display bfd session static** [**for-ip** | **for-lsp** | **for-te**] [**verbose**]：查看静态 BFD 相关配置信息。
- **display bfd session peer-ip** *peer-ip* [**vpn-instance** *vpn-instance-name*] [**verbose**]：查看反向通道为 IP 的 BFD 配置信息
- **display bfd session static-lsp** *lsp-name* [**verbose**]：查看反向通道为静态 LSP 的 BFD 配置信息。
- **display bfd session ldp-lsp peer-ip** *peer-ip* **nexthop** *nexthop-address* [**interface**

interface-type interface-number] [**verbose**]：查看反向通道为 LDP LSP 的 BFD 配置信息。

- **display bfd session mpls-te interface tunnel** *interface-number* **te-lsp** [**verbose**]：查看反向通道为 CR-LSP 的 BFD 配置信息。
- **display bfd session mpls-te interface tunnel** *interface-number* [**verbose**]：查看反向通道为 TE 隧道的 BFD 配置信息。

■ 执行以下命令查看 BFD 统计信息。

- **display bfd statistics session all** [**for-ip** | **for-lsp** | **for-te**]：查看所有 BFD 会话的统计信息。
- **display bfd statistics session peer-ip** *peer-ip* [**vpn-instance** *vpn-instance-name*]：查看检测 IP 链路的 BFD 会话统计信息。
- **display bfd statistics session static-lsp** *lsp-name*：查看检测静态 LSP 的 BFD 会话统计信息。
- **display bfd statistics session ldp-lsp peer-ip** *peer-ip* **nexthop** *nexthop-address* [**interface** *interface-type interface-number*]：查看检测 LDP LSP 的 BFD 会话统计信息。
- **display bfd statistics session mpls-te interface tunnel** *interface-number* **te-lsp**：查看检测 CR-LSP 的 BFD 会话统计信息。
- **display bfd statistics session mpls-te interface tunnel** *interface-number*：查看检测 TE 隧道的 BFD 会话统计信息。

已经完成动态 BFD for CR-LSP 功能的所有配置后，执行以下 **display** 命令查看相关配置，验证配置效果。

■ **display bfd configuration dynamic** [**verbose**]：在隧道入节点上查看动态 BFD 检测的配置信息。

■ **display bfd configuration passive-dynamic** [**peer-ip** *peer-ip* **remote-discriminator** *discriminator*] [**verbose**]：在隧道出节点上查看动态 BFD 检测配置信息。

■ **display bfd session dynamic** [**verbose**]：在隧道入节点上查看动态 BFD 会话信息。

■ **display bfd session passive-dynamic** [**peer-ip** *peer-ip* **remote-discriminator** *remote-discr-value*] [**verbose**]：在隧道出节点上查看被动创建的 BFD 会话信息。

■ 执行命令以下，查看 BFD 统计信息。

- **display bfd statistics**：查看 BFD 所有相关统计信息。
- **display bfd statistics session dynamic**：查看动态 BFD 会话相关统计信息。

■ **display mpls bfd session** [**fec** *fec-address* | **monitor** | **nexthop** *ip-address* | **outgoing-interface** *interface-type interface-number* | **statistics** | **verbose**]或 **display mpls bfd session protocol** { **cr-static** | **rsvp-te** } [**lsp-id** *ingress-lsr-id session-id lsp-id* [**verbose**]]：查看与 MPLS 相关的 BFD 会话信息。

7.2.6　静态 BFD for CR-LSP 配置示例

在如图 7-5 所示的 MPLS 网络中，要求从 LSRA 上建立一条 TE 隧道，目的地址为 LSRC，并配置 CR-LSP 热备份（hot-standby）和逃生路径（best-effort）。其中路径如下。

■ 主 CR-LSP 的路径为 LSRA→LSRB→LSRC。

■ 备份 CR-LSP 的路径为 LSRA→LSRD→LSRC。

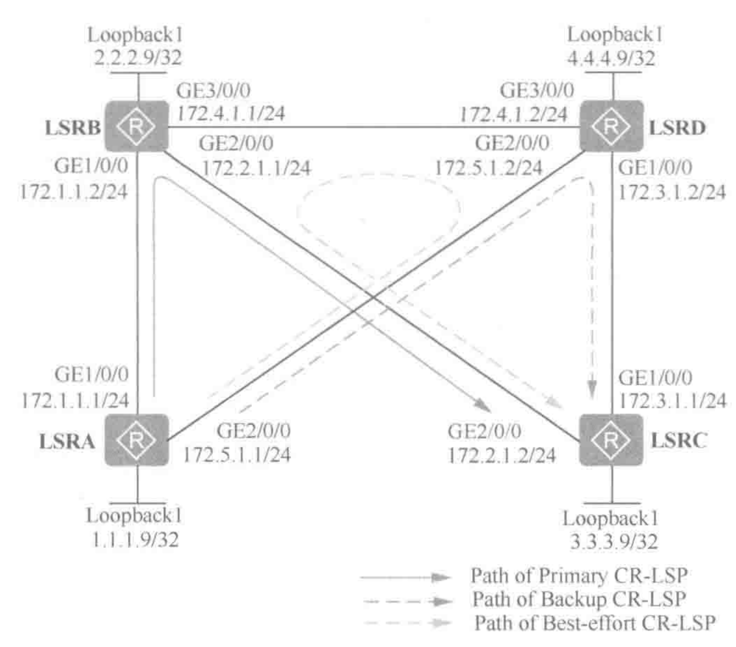


图 7-5 静态 BFD for CR-LSP 配置示例的拓扑结构

当主 CR-LSP 故障时，流量切换到备份 CR-LSP；当主 CR-LSP 故障恢复，延时 15s 后再进行流量回切。如果主、备 CR-LSP 都发生故障，触发建立逃生路径，使流量切换到逃生路径上。其中主、备 CR-LSP 都可以通过显式路径来指定，逃生路径由系统自动根据网络的故障情况来计算不需用户指定，本例中的结果为 LSRA→LSRD→LSRB→LSRC。当故障节点不同时，逃生路径的结果也不一样。

现要求配置两个静态 BFD 会话，分别检测主、备 CR-LSP，使得：

■ 主 CR-LSP 故障时，将流量快速切换到备份 CR-LSP；

■ 在主 CR-LSP 恢复后的延时回切时间（15s）内，如果备份 CR-LSP 故障，可快速感知故障并将流量回切到主 CR-LSP。

1. 基本配置思路分析

本示例与 7.1.8 节介绍的配置示例的拓扑结构和各接口 IP 地址完全一样，在主 CR-LSP、备份 CR-LSP 和逃生路径方面的配置要求也完全一样，唯一不同的本示例要配置针对主、备 CR-LSP 的静态 BFD 检测。故本示例的基本配置思路如下。

（1）在各节点上配置各接口（包括 Loopback 接口）的 IP 地址和 OSPF 协议，实现各节点之间公网路由可达。

（2）在各节点上配置 LSR ID，并使能各节点全局和公网接口的 MPLS、MPLS TE 和 RSVP-TE 能力，在入节点 LSRA 上使能 CSPF 功能。本示例仅介绍从 LSRA 到 LSRC 的单向 TE 隧道的配置，故入节点仅为 LSRA，实际应用中需要配置双向 TE 隧道，也就要进行双向 TE 隧道的对应配置了。

（3）在各节点上使能 OSPF TE，使得 OSPF 协议可以发布 MPLS TE 信息。

（4）在入节点 LSRA 上配置主、备 CR-LSP 的显式路径。

（5）在入节点 LSRA 上创建目的地址为 LSRC 的隧道接口，指定前面配置的主备 CR-LSP 的显式路径，并使能热备份和逃生路径，配置回切时间为 15s。

（6）在入节点 LSRA 建立两个 BFD 会话，分别检测主 CR-LSP 和备份 CR-LSP；出节点上配置两个 BFD 会话，绑定 IP 链路（LSRC→LSRA 路由可达即可）。

2. 具体配置步骤

以上配置任务中的第（1）~（5）项的具体配置与 7.1.8 节示例中对应的配置完全一样，参见即可。下面仅介绍以上第（6）项配置任务的配置方法。

（6）在入节点 LSRA 和出节点 LSRC 上配置静态 BFD for CR-LSP。

在 LSRA 和 LSRC 之间建立 BFD 会话检测主、备 CR-LSP 故障。LSRA 上的 BFD 会话绑定 CR-LSP，LSRC 上的 BFD 会话绑定 IP 链路。指定发送 BFD 报文的时间间隔和允许接收 BFD 报文的时间间隔为 500ms，BFD 本地检测倍数为 3。

LSRA 上的配置。

```
[LSRA] bfd
[LSRA-bfd] quit
[LSRA] bfd prilsp2lsrc bind mpls-te interface tunnel 0/0/1 te-lsp   #---创建主 CR-LSP 的 BFD 会话
[LSRA-bfd-lsp-session-prilsp2lsrc] discriminator local 139
[LSRA-bfd-lsp-session-prilsp2lsrc] discriminator remote 239
[LSRA-bfd-lsp-session-prilsp2lsrc] min-tx-interval 500
[LSRA-bfd-lsp-session-prilsp2lsrc] min-rx-interval 500
[LSRA-bfd-lsp-session-prilsp2lsrc] detect-multiplier 3
[LSRA-bfd-lsp-session-prilsp2lsrc] process-pst
[LSRA-bfd-lsp-session-prilsp2lsrc] notify neighbor-down
[LSRA-bfd-lsp-session-prilsp2lsrc] commit
[LSRA-bfd-lsp-session-prilsp2lsrc] quit
[LSRA] bfd backuplsp2lsrc bind mpls-te interface tunnel 0/0/1 te-lsp backup   #--创建备份 CR-LSP 的 BFD 会话
[LSRA-bfd-lsp-session-backuplsp2lsrc] discriminator local 339
[LSRA-bfd-lsp-session-backuplsp2lsrc] discriminator remote 439
[LSRA-bfd-lsp-session-backuplsp2lsrc] min-tx-interval 500
[LSRA-bfd-lsp-session-backuplsp2lsrc] min-rx-interval 500
[LSRA-bfd-lsp-session-backuplsp2lsrc] detect-multiplier 3
[LSRA-bfd-lsp-session-backuplsp2lsrc] process-pst
[LSRA-bfd-lsp-session-backuplsp2lsrc] notify neighbor-down
[LSRA-bfd-lsp-session-backuplsp2lsrc] commit
[LSRA-bfd-lsp-session-backuplsp2lsrc] quit
```

LSRC 上的配置。

```
[LSRC] bfd
[LSRC-bfd] quit
[LSRC] bfd reversepri2lsra bind peer-ip 1.1.1.9
[LSRC-bfd-session-reversepri2lsra] discriminator local 239
[LSRC-bfd-session-reversepri2lsra] discriminator remote 139
[LSRC-bfd-session-reversepri2lsra] min-tx-interval 500
[LSRC-bfd-session-reversepri2lsra] min-rx-interval 500
[LSRC-bfd-session-reversepri2lsra] detect-multiplier 3
[LSRC-bfd-session-reversepri2lsra] commit
[LSRC-bfd-session-reversepri2lsra] quit
[LSRC] bfd reversebac2lsra bind peer-ip 1.1.1.9
[LSRC-bfd-session-reversebac2lsra] discriminator local 439
[LSRC-bfd-session-reversebac2lsra] discriminator remote 339
[LSRC-bfd-session-reversebac2lsra] min-tx-interval 500
[LSRC-bfd-session-reversebac2lsra] min-rx-interval 500
```

```
[LSRC-bfd-session-reversebac2lsra] detect-multiplier 3
[LSRC-bfd-session-reversebac2lsra] commit
[LSRC-bfd-session-reversebac2lsra] quit
```

以上配置完成后，在 LSRA 和 LSRC 上执行 **display bfd session discriminator** 命令，可发现 BFD 会话状态为 Up。

以 LSRA 的显示为例：

```
[LSRA] display bfd session discriminator 139
--------------------------------------------------------------------------------
Local Remote     PeerIpAddr      State     Type         InterfaceName
--------------------------------------------------------------------------------
139   239        3.3.3.9         Up        S_TE_LSP     Tunnel0/0/1
--------------------------------------------------------------------------------
[LSRA] display bfd session discriminator 339
--------------------------------------------------------------------------------
Local Remote     PeerIpAddr      State     Type         InterfaceName
--------------------------------------------------------------------------------
339   439        3.3.3.9         Up        S_TE_LSP     Tunnel0/0/1
--------------------------------------------------------------------------------
```

3. 配置结果验证

为了模拟在主 CR-LSP 恢复后的延时回切时间（15s）内备份 CR-LSP 故障的场景，可以通过在 GE1/0/0 接口插入线缆后，在 LSRA 重复执行 **display mpls te tunnel-interface tunnel** 0/0/1 命令查看隧道信息，直到发现主 CR-LSP 建立成功，然后在 15s 内拔出 LSRA 或 LSRD 的 GE2/0/0 接口线缆来实现。此时可发现流量快速回切到主 CR-LSP，故障收敛时间为毫秒级。证明所配置的静态 BFD 检测功能正常起作用了。

7.2.7 动态 BFD for CR-LSP 配置示例

在如图 7-6 所示的 MPLS 网络中，要求从 LSRA 上建立一条 TE 隧道，目的地址为 LSRC，并配置 CR-LSP 热备份（hot-standby）和逃生路径（best-effort）。其中：

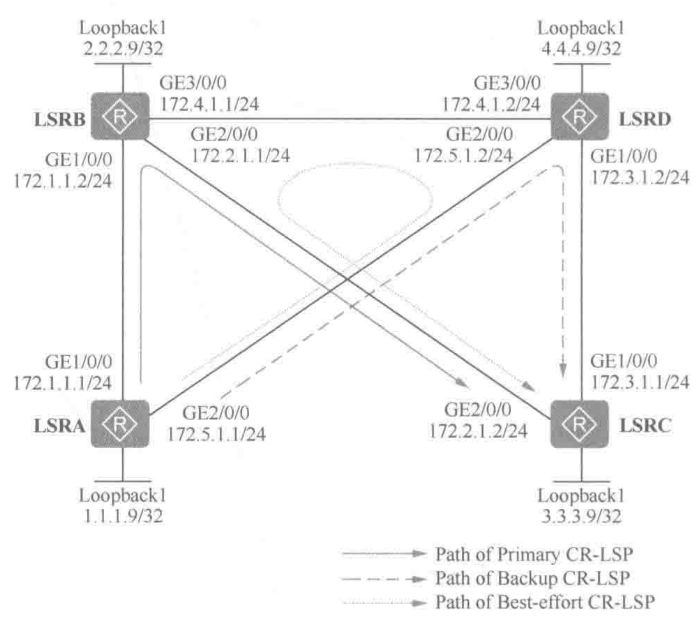


图 7-6 动态 BFD for CR-LSP 配置示例的拓扑结构

■ 主 CR-LSP 的路径为 LSRA→LSRB→LSRC；

■ 备份 CR-LSP 的路径为 LSRA→LSRD→LSRC。

当主 CR-LSP 故障时，流量切换到备份 CR-LSP；当主 CR-LSP 故障恢复，延时 15s 后再进行流量回切。如果主、备 CR-LSP 都故障，触发建立逃生路径，使流量切换到逃生路径上。其中主、备 CR-LSP 都可以通过显式路径来指定，逃生路径由系统自动根据网络的故障情况来计算不需用户指定，本例中的结果为 LSRA→LSRD→LSRB→LSRC。当故障节点不同时，逃生路径的结果也不一样。

现要求配置动态 BFD for CR-LSP 检测主、备 CR-LSP，使得：

■ 主 CR-LSP 故障时，流量快速切换到备份 CR-LSP；

■ 在主 CR-LSP 恢复后的延时回切时间（15s）内，如果备份 CR-LSP 故障，可快速感知故障并将流量回切到主 CR-LSP。

1. 基本配置思路分析

本示例的拓扑结构、各接口 IP 地址配置与 7.1.8 节介绍的示例完全一样，在配置要求上也仅多了一个针对主、备 CR-LSP 建立的动态 BFD 会话功能。故本示例的基本配置思路如下。

（1）在各节点上配置各接口（包括 Loopback 接口）的 IP 地址和 OSPF 协议，实现各节点之间公网路由可达。

（2）在各节点上配置 LSR ID，并使能各节点全局和公网接口的 MPLS、MPLS TE 和 RSVP-TE 能力，在入节点 LSRA 上使能 CSPF 功能。本示例仅介绍从 LSRA 到 LSRC 的单向 TE 隧道的配置，故入节点仅为 LSRA，实际应用中需要配置双向 TE 隧道，也就要进行双向 TE 隧道的对应配置了。

（3）在各节点上使能 OSPF TE，使得 OSPF 协议可以发布 MPLS TE 信息。

（4）在入节点 LSRA 上配置主、备 CR-LSP 的显式路径。

（5）在入节点 LSRA 上创建目的地址为 LSRC 的隧道接口，指定前面配置的主备 CR-LSP 的显式路径，并使能热备份和逃生路径，配置回切时间为 15s。

（6）在入节点使能 BFD，配置动态 BFD for CR-LSP 功能，指定本地发送 BFD 报文的时间间隔和允许接收 BFD 报文的时间间隔，以及 BFD 本地检测倍数。在出节点使能被动创建 BFD 会话。

2. 具体配置步骤

以上配置任务中的第（1）～（5）项的具体配置与 7.1.8 节示例中对应的配置完全一样，参见即可。下面仅介绍以上第（6）项配置任务的配置方法。

（6）在入节点配置动态 BFD for CR-LSP，在出节点使能被动创建 BFD 会话。

在入节点配置动态 BFD for CR-LSP 功能，指定本地发送 BFD 报文的时间间隔和允许接收的时间间隔为 500ms，BFD 本地检测倍数为 3。

LSRA 上的配置。

```
[LSRA] bfd
[LSRA-bfd] quit
[LSRA] interface tunnel 0/0/1
[LSRA-Tunnel0/0/1] mpls te bfd enable
[LSRA-Tunnel0/0/1] mpls te bfd min-tx-interval 500 min-rx-interval 500 detect-multiplier 3
[LSRA-Tunnel0/0/1] mpls te commit
```

\# LSRC 上的配置。

```
[LSRC] bfd
[LSRC-bfd] mpls-passive
[LSRC-bfd] quit
```

以上配置完成后，在 LSRA 上执行 **display bfd session mpls-te interface Tunnel 0/0/1 te-lsp** 命令，可发现动态 BFD 会话状态为 Up，BFD 会话类型为 D_TE_LSP（动态创建且与 TE-LSP 绑定）。

```
[LSRA] display bfd session mpls-te interface Tunnel 0/0/1 te-lsp
--------------------------------------------------------------------------------
Local Remote     PeerIpAddr      State     Type        InterfaceName
--------------------------------------------------------------------------------
8192   8192       3.3.3.9        Up        D_TE_LSP     Tunnel0/0/1
--------------------------------------------------------------------------------
     Total Up/DOWN Session Number : 1/0
```

在 LSRC 上执行 **display bfd session passive-dynamic** 命令，从创建的 BFD 会话类型为 E_ Dynamic（完全动态会话，使能被动动态创建 BFD 会话功能后所创建的 BFD 会话类型）可发现被动建立了一个 BFD 会话。

```
[LSRC] display bfd session passive-dynamic
--------------------------------------------------------------------------------
Local Remote     PeerIpAddr      State     Type        InterfaceName
--------------------------------------------------------------------------------
8192   8192       1.1.1.9        Up        E_Dynamic       -
--------------------------------------------------------------------------------
     Total Up/DOWN Session Number : 1/0
```

3. 配置结果验证

为了模拟在主 CR-LSP 恢复后的延时回切时间（15s）内备份 CR-LSP 故障的场景，可以通过在 GE1/0/0 接口插入线缆后，在 LSRA 上重复执行 **display mpls te tunnel-interface tunnel** 0/0/1 命令查看隧道信息，直到发现主 CR-LSP 建立成功，然后在 15s 内拔出 LSRA 或 LSRD 的 GE2/0/0 接口线缆来实现。此时可发现流量快速回切到主 CR-LSP，故障收敛时间为毫秒级。证明所创建的动态 BFD 会话正常起作用了。



第8章 MPLS QoS配置与管理



前面各章详细介绍了与普通的 MPLS 隧道和 MPLS TE 隧道建立相关的技术原理和各项功能的配置管理方法。本章以及第 9 章要专门介绍 MPLS 隧道、MPLS TE 隧道的 QoS 服务功能的配置与管理方法。

本章所介绍的 MPLS QoS 服务功能配置与管理方法是专门针对华为 S 系列交换机的，相对比较简单。S 系列交换机中的 MPLS QoS 服务功能是通过扩展 IP 网络 QoS 中的 DiffServ 服务模型，配置 MPLS 报头中的 EXP 优先级与 IP 报头中的 DSCP 优先级、802.1p 优先级，以及所映射的 PHB（下一跳行为）、LP（本地优先级）优先级之间的映射关系而实现的。第 9 章所介绍的 MPLS DS-TE 是华为 AR G3 系列路由器的 MPLS QoS 功能实现方式，要复杂许多。

8.1 MPLS QoS 基础

在 MPLS QoS（Quality of Service，服务质量）功能方面，华为 S 系列交换机比较简单，是把 MPLS 与 IP 网络中的 DiffServ 模型结合起来，通过将 IP 报文中的 DS（Differentiated Service，差分服务）字段值或 VLAN 报文中的 PRI 字段值的分配与 MPLS 的标签分配过程结合，利用 MPLS 中的 EXP 优先级与 DiffServ 优先级（如 DSCP 优先级、802.1p 优先级）之间的映射来实现为不同类型流量提供不同服务等级。

在 AR G3 系列路由器中，MPLS QoS 功能的实现要复杂许多，也要强大许多，它是通过结合了 MPLS TE 和 MPLS DiffServ 的 MPLS DS-TE（DiffServ-aware Traffic Engineering，差分感知流量工程）功能实现的，不仅可以为不同类型流量提供不同的服务等级，还可以为不同类型流量配置不同的预留带宽保证服务，因为在 MPLS TE 中有用于资源预留协议——RSVP-TE。

本章先来介绍华为 S 系列交换机中的 MPLS QoS 功能技术原理及配置与管理方法。

说明 S 系列交换机仅 S5700 系列及更高配置系列（SI 和 LI 子系列除外）的机型支持 MPLS QoS 功能配置，具体请参见相应的产品手册说明。

8.1.1 MPLS DiffServ 简介

目前 MPLS 被广泛地应用于大规模网络的组建，而在 MPLS 网络中，因为三层协议报头已成为了 MPLS 报文的数据部分，无法通过 IP 报头的优先级值来实现服务质量，所以在 MPLS 网络中的服务质量就需要采用其他方式来实现，那就是此处介绍的 MPLS QoS。

与 IP 报文根据报文中携带的 IP 网络优先级（如 DSCP 优先级、802.1p 优先级）来区分业务的服务等级类似，MPLS QoS 是根据 MPLS 标签中的 EXP 优先级来区分不同的数据流，实现差分服务，保证语音、视频数据流的低延时、低丢包率，保证网络的高利用率。

MPLS QoS 与 IP QoS 一样，也是通过差分服务（DiffServ）模型来实施 QoS 的，但

在MPLS QoS中的DiffServ是结合了MPLS标签转发和DiffServ差分服务两项功能，所以称之为MPLS DiffServ。通过MPLS DiffServ可以为每个经MPLS网络传输的业务提供特定的服务，并提供差异化的服务类型来满足各种需求。

DiffServ的基本机制是在IP网络边缘，利用IP报文中的DSCP优先级或VLAN报文中的802.1p优先级唯一地标记该业务分类，然后骨干网络中的各LSR（P节点）根据该字段对各种业务采取预先设定的服务策略，保证相应的服务质量。华为S系列交换机中的MPLS QoS功能是基于MPLS DiffServ来实现的，它通过将MPLS报文中的EXP优先级映射到DiffServ PHB（Per-Hop Behavior，下一跳行为），然后骨干网中各LSR可根据MPLS报文的EXP优先级指导报文转发。

MPLS DiffServ提供了两种实现方案。

■ E-LSP路径，即由MPLS报文中的EXP优先级决定PHB的LSP。

该方案适用于支持少于8个PHB的网络（因为EXP只有8个优先级值），将指定的DSCP或802.1p优先级映射为特定的内层MPLS标签（在同一MPLS隧道中传输的报文其外层公网隧道标签是相同的）EXP优先级，标识到特定的PHB。在转发过程中，报文根据MPLS隧道标签转发，而由EXP优先级决定在每一跳LSR上的调度和丢弃优先级，因此同一条LSP可以承载8类不同PHB的流，通过MPLS报头的EXP字段值来进行区分。EXP优先级可以直接由用户配置决定，也可以从报文的DSCP或802.1p优先级直接映射得到。PHB与EXP的缺省映射关系如表8-1所示。这种方法不需要信令协议传递PHB信息，而且标签使用率较高，状态易于维护。

表8-1　　PHB与EXP的缺省映射关系

PHB行为	EXP优先级
BE	0
AF1	1
AF2	2
AF3	3
AF4	4
EF	5
CS6	6
CS7	7

■ L-LSP路径，即由MPLS标签和EXP优先级共同决定PHB的LSP。

该方案适用于支持任意数量PHB的网络，因为在这种方案中，MPLS隧道标签不仅用于决定转发路径，而且决定在LSR上的调度行为（带有不同MPLS标签的流量进入不同的LSP转发），而EXP优先级则用于决定数据报文的丢弃优先级。由于要通过MPLS标签来区分业务流的类型，所以需要为不同的流建立不同的LSP，需要使用更多的标签，占用大量的系统资源。目前暂不支持L-LSP方案。

8.1.2 Diffserv域

如图8-1所示，在整个MPLS网络的Diffserv域中，可以分成MPLS Diffserv域和IP Diffserv域两个部分，分别用于MPLS报文的转发和IP报文的转发。

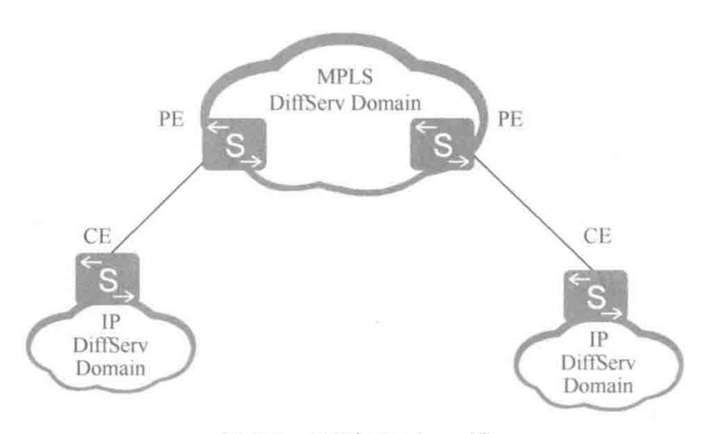


图 8-1 两种 Diffserv 域

在上节介绍的 E-LSP 方案中，在 MPLS 网络的边缘，MPLS Diffserv 在两个 Diffserv 域之间进行协调管理调度，完成 DSCP 或 802.1p 优先级到 EXP 优先级的双向映射。

在 E-LSP 方案中，MPLS 报文基于 EXP 优先级采用相应的转发行为，为客户提供不同的服务质量。如图 8-2 所示，在 MPLS 报文进入 P 节点时需要进行流分类，将报文携带的 EXP 优先级（在 PE_1 节点上已配置好）统一映射到设备内部的对应服务等级（对应不同的队列）和丢弃优先级。流量分类后，流量整形、流量监管、拥塞避免等 QoS 实现方法就和 IP 网络中的完全相同了。在报文从 P 节点发出时，再将内部的服务等级和丢弃优先级映射为 EXP 优先级，以便后续网络设备根据报文中的 EXP 优先级提供相应的服务质量。

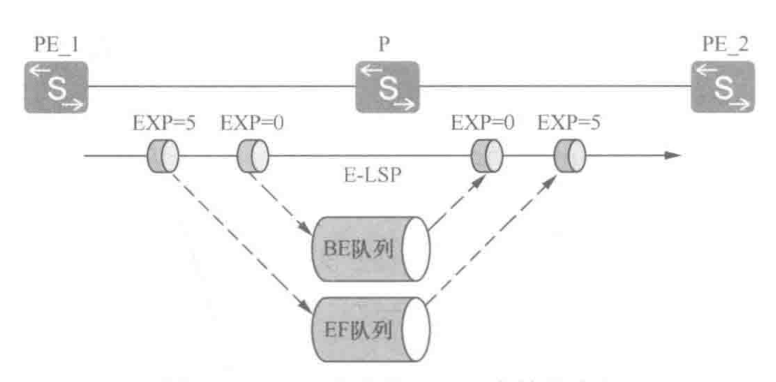


图 8-2 E-LSP 方案的报文基本转发流程

8.1.3 MPLS DiffServ 的工作模式

MPLS DiffServ 有三种工作模式：Uniform、Pipe 和 Short pipe，下面分别予以介绍。

1. Uniform 模式

在 Uniform（统一）工作模式中，IP 网络中的 IP 优先级和 MPLS 网络中的 EXP 优先级之间是相互联动的，具体特征体现如下。

■ 在 Ingress 节点上，报文在被打上 MPLS 标签后，根据报本地设备配置的 DSCP 或 802.1p 优先级与 EXP 优先级的映射关系，将 IP 报文中的 DSCP 或 802.1p 字段值映射到新增的每个 MPLS 标签中的 EXP 字段中。

■ 外层 MPLS 标签在弹出时会先自动把该标签中的 EXP 优先级值复制到新的外层 MPLS 标签中的 EXP 字段上。

■ 如果报文在 MPLS 网络传输过程中某 MPLS 标签的 EXP 字段值改变了，到了 Egress 节点时可能会影响报文在离开 MPLS 网络后采用的 PHB。因为在 Egress 节点会根据本地配置的 DSCP 或 802.1p 优先级与 EXP 优先级的映射关系，反向将当前外层 MPLS 标签中的 EXP 优先级值映射到 IP 报文中的 DSCP 或 802.1p 优先级值。

当电信运营商认为可以完全信任 CE 侧流量携带过来的 QoS 参数时，可以采用 Uniform 模式。下面以如图 8-3 所示的 L3VPN 为例介绍 Uniform 模式基本工作原理。

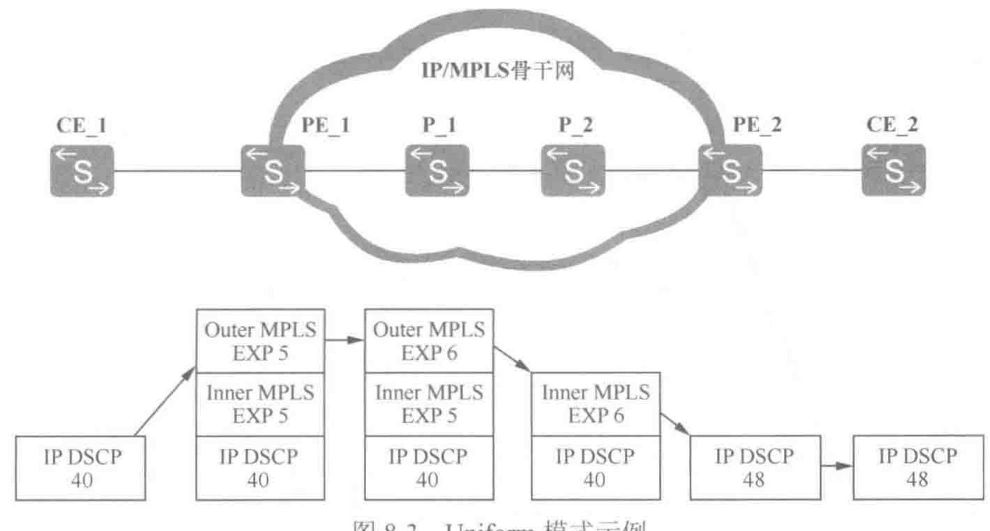


图 8-3　Uniform 模式示例

（1）在 PE_1 上根据本地配置的 DSCP 优先级与 EXP 优先级的映射关系（此处假设采用缺省映射关系），将 IP 报文中携带的 DSCP 优先级 40 映射到新增的两层 MPLS 标签中的 EXP 字段中，使它们的 EXP 优先级均为 5。

（2）假设 MPLS 报文传输到 P_1 时将 Outer MPLS 标签（为公网隧道标签）的 EXP 值改为 6，然后继续传输到 P_2，在 P_2 上保持不变。

（3）在 P_2 继续向 Egress 节点传输 MPLS 报文时，会先弹出 Outer MPLS 标签，但在弹出前会将 Outer MPLS 标签中的 EXP 优先级值 6 复制到新的外层 MPLS 标签——Inner MPLS 标签（在 L3VPN 中为私网路由标签，在 L2VPN 中为私网 VC 标签）的 EXP 字段上，这样 Inner MPLS 标签的 EXP 优先级值就由原来的 5 变为 6 了。

（4）因为 PE_2 是 Egress 节点，所以在继续向 IP 网络传输报文前，必须弹出剩下的 Inner MPLS 标签。但在弹出前又要根据本地设备上配置的 EXP 优先级与 DSCP 优先级之间的映射关系（本示例也假设采用缺省映射关系），修改发送的 IP 报文的 DSCP 优先级值为对应的 48。

通过以上步骤，就可以使得源端 CE 发送的 IP 报文中的 DSCP 优先级在经过 MPLS 网络传输时会随着 MPLS 标签中的 EXP 优先级的变化而变化，到达目的端 CE 时 IP 报文中所携带的 IP 网络优先级值可能与源端的不一样了。从中可以看出，EXP 优先级的变化将带来全局的报文转发影响，而不是仅限于 MPLS 网络中的报文转发。

说明　本示例以及下面两个示例均以非 PHP 场景的 L3VPN 为例进行介绍。在 PHP 场景下，PE_2 设备收到的报文不包含 MPLS EXP 信息，因为 MPLS 标签已在倒数第二跳弹出了，也就不会进行 MPLS QoS 处理了。如果要进行 MPLS QoS，建议将原来计划部署在出节点上的相关配置部署到倒数第二跳设备上，或者在出节点部署向倒数第二跳分配显式空标签（IPv4 环境为 0，IPv6 环境为 2），这样标签就不会在倒数第二跳弹出了。

2. Pipe 模式

Pipe（管道）模式可以这么理解：报文从 Ingress 节点进入 MPLS 网络，到 Egress 节点离开 MPLS 网络，报文中所携带的 IP 网络优先级不变。但在这种模式中，从 Egress 节点发送到目的站点 CE 时的 PHB 仍由最后一个被弹出的 MPSL 标签中的 EXP 优先级决定。具体特征体现如下。

- IP 报文进入 Ingress 节点后，在新增 MPLS 标签中的 EXP 值可以由用户指定。
- 如果报文在 MPLS 网络传输中改变了某个 MPLS 标签的 EXP 字段值，则仅对报文在 MPLS 网络内传输有影响。
- 在 Egress 节点上，报文会根据外层 MPLS 标签中的 EXP 字段值选择 PHB，而当报文离开 MPLS 网络后，直接使用原 IP 报文中携带的 DSCP 或 802.1p。

当电信运营商完全不关心 CE 侧用户设置的 QoS 参数时，在 PE 上为新增 MPLS 标签手动指定 EXP 字段值，不受 IP 报文中携带的 IP 网络优先级影响。这样一来，从 Ingress 节点到 Egress 节点都是按照运营商的意愿在各 P 节点进行 QoS 调度，直到将流量送出最后一个 P 节点之后，报文再根据其原来携带的 IP 网络优先级值转发。

下面以如图 8-4 所示的非 PHP 场景 L3VPN 为例介绍 Pipe 模式基本工作原理。

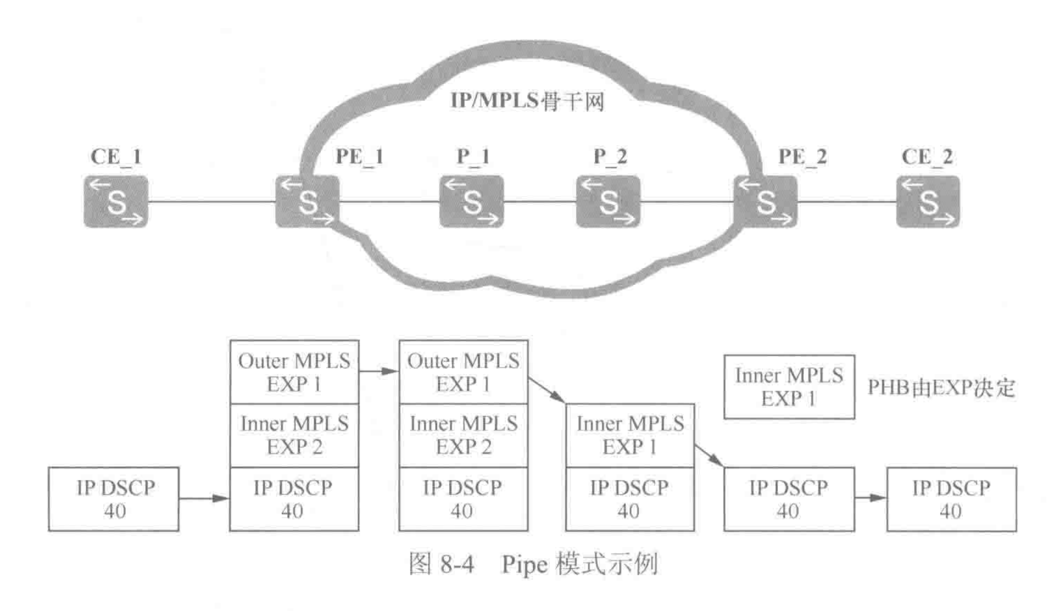


图 8-4　Pipe 模式示例

（1）当 IP 报文进入 Ingress 节点 PE_1 后，直接根据本地配置为新增的 Outer MPLS 和 Inner MPLS 标签中的 EXP 赋值，假设分别为 1 和 2。

（2）当 MPLS 报文传输到 P_1 节点时，假设对 Outer MPLS 标签中的 EXP 优先级进行了修改，由原来的 1 改为 2。但这个 EXP 优先级的修改不会影响报文原来携带的 DSCP

优先级值。

（3）P_1 按照 Outer MPLS 标签中新的 EXP 优先值进行报文转发，到了 P_2 节点时 Outer MPLS 标签要被弹出，此时也会先将 Outer MPLS 标签中的 EXP 优先级值复制到新外层 MPLS 标签——Inner MPLS 标签中的 EXP 字段上，即 Inner MPLS 标签中的 EXP 优先级值由原来的 2 变为 1。

（4）MPLS 报文继续转发到 Egress 节点 PE_2 时，Inner MPLS 标签也将被弹出，但在弹出前仍旧会根据 Inner MPLS 标签中的 EXP 优先级值查找对应映射的 PHB，然后在转发到 CE_2 时再根据 IP 报文中携带的 DSCP 优先级值进行转发。

从以上流程可以看出，源端 CE 发送的 IP 报文中携带的 IP 网络优先级值，通过 MPLS 网络传输后，到达目的端 CE 时 IP 报文中所携带的 IP 网络优先级值会保持不变。

3. Short pipe 模式

Short pipe（短管道）模式是对 Pipe 模式的改进型，报文在进入 P 节点时对报文的处理方式与 Pipe 模式相同，区别仅是在 Egress 节点的倒数第二跳，就完成了 IP 报文中的 QoS 参数恢复。也就是，从 Ingress 节点到倒数第二跳的 P 节点，与 Pipe 模式一样，全部是按照运营商的意愿进行 QoS 调度，只是到了 Egress 节点时不用再由 MPLS 标签的 EXP 优先级来决定 PHB，而是直接根据 IP 报文中所携带的 IP 网络优先级决定 PHB 了。由此可以看出，Short pipe 模式中，EXP 优先级值对报文转发的影响比 Pipe 模式还要小。

下面以如图 8-5 所示的非 PHP 场景 L3VPN 为例介绍 Short pipe 模式的基本工作原理。前面 3 步与 Pipe 模式中的前面 3 步完全一样，只是到了第（4）步时，PE_2 会先将 Inner MPLS 标签弹出，然后直接根据 IP 报文中的 DSCP 优先级值选择 PHB（而不用根据 Inner MPLS 标签中的 EXP 优先级值选择 PHB），然后转发给目的端 CE_2。

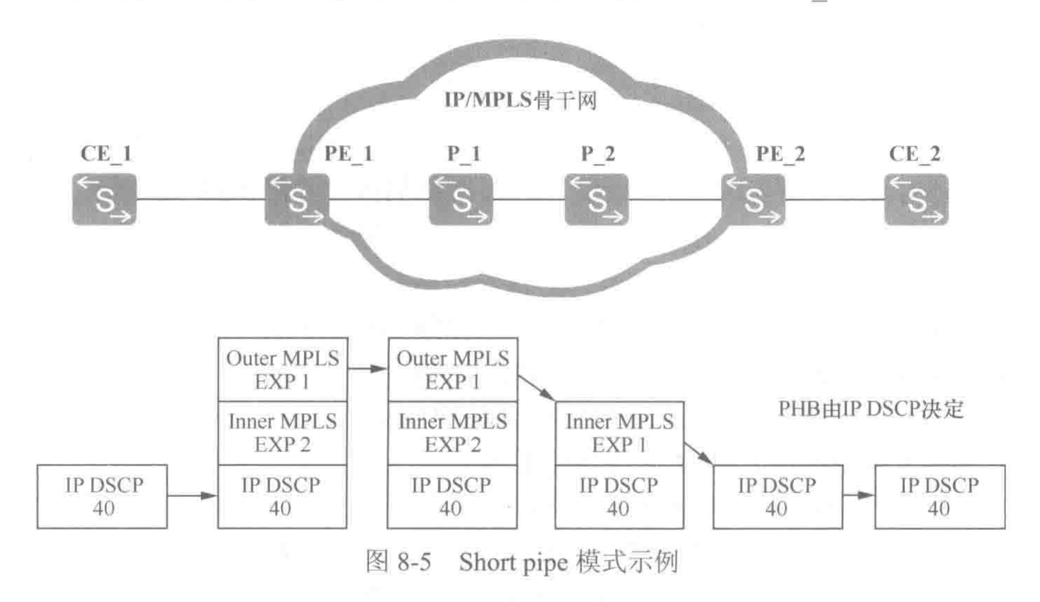


图 8-5　Short pipe 模式示例

8.1.4　MPLS QoS 在 VPN 业务中的应用

随着 MPLS 技术得到广泛的应用，很多服务提供商通过 MPLS 网络向企业提供 VPN 业务。VPN 可以用于连接出差人员与企业总部、异地分支机构与企业总部、企业合作伙

件与企业总部等，提供它们之间的信息传输。但是如果 VPN 不能保证企业运营数据的及时有效发送，那么 VPN 将仍然不能有效地为企业服务。例如，企业内语音、视频需要受到优先对待，保证这些应用的带宽要求，而对于 E-Mail、WWW 访问等则可以尽最大的可能性来发送，但对时延、可靠性等性能不提供任何保证。

为了满足企业 VPN 的需求，可以通过部署 MPLS QoS 来实现以下两种应用。

1. 区分 VPN 内不同业务的优先级

如图 8-6 所示，这两个 Site 属于同一企业的不同分部，企业网络中存在语音、数据和视频等多种业务流，已在企业网络中对三类业务区分优先级，保证语音优先级最高、视频其次、数据优先级最低。当不同 VPN 业务流量进入 MPLS 网络时，需要在 MPLS 网络中对三类业务区分优先级，保证 MPLS 网络中语音优先级一直最高、视频其次、数据优先级最低，并根据优先级的高低对三类业务提供不同的 QoS 服务。

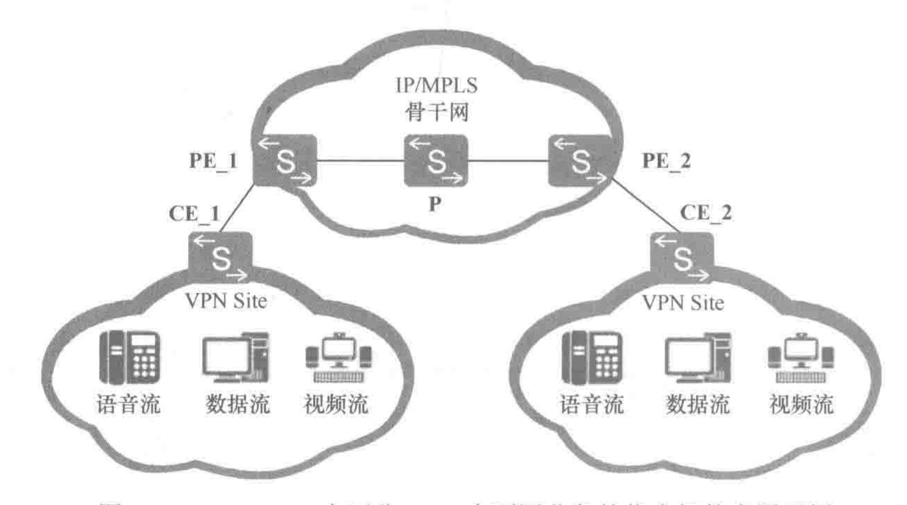


图 8-6 MPLS QoS 在区分 VPN 内不同业务的优先级的应用示例

不同网络中的报文使用不同的优先级字段，例如二层网络的报文使用 802.1p 优先级，三层网络的报文使用 DSCP 优先级，MPLS 网络的报文使用 EXP 优先级。下面以图 8-6 中 L3VPN 报文从 PE_1 发往 PE_2 的过程为例介绍 MPLS QoS 在区分同一 VPN 网络内部不同业务的优先级方面的应用。

（1）入节点（PE_1）将根据本地配置，接收到的 IP 报文携带的 DSCP 优先级映射到内部服务等级和颜色，再根据服务等级和颜色对报文进行不同的 QoS 服务。报文在流出设备时，设备又会根据内部服务等级和颜色标记为新增的MPLS标签中的EXP优先级，以便后续 MPLS 网络根据 EXP 优先级进行服务。

（2）中间节点（P）根据本地配置，将接收到的报文携带的 EXP 优先级映射到内部服务等级和颜色，再根据服务等级和颜色对报文进行不同的 QoS 服务。报文在流出设备时，设备根据内部服务等级和颜色重标记为 EXP 优先级。

（3）出节点（PE_2）根据本地配置，将接收到的报文携带的 EXP 或 DSCP 优先级（具体要视 MPLS DiffServ 工作模式，以及倒数第二跳分配的 MPLS 标签是否支持 PHP 而定）映射到内部服务等级和颜色，再根据服务等级和颜色对报文进行不同的 QoS 服务，设备再根据内部服务等级和颜色重标记为 IP 报文的 DSCP 优先级，以便后续网络根据报

文优先级进行服务。

2. 区分不同 VPN 的优先级

如图 8-7 所示，CE_1 和 CE_3 属于 VPN_1，并分别连接企业 A 的两个分部；CE_2 和 CE_4 属于 VPN_2，并分别连接企业 B 的两个分部。

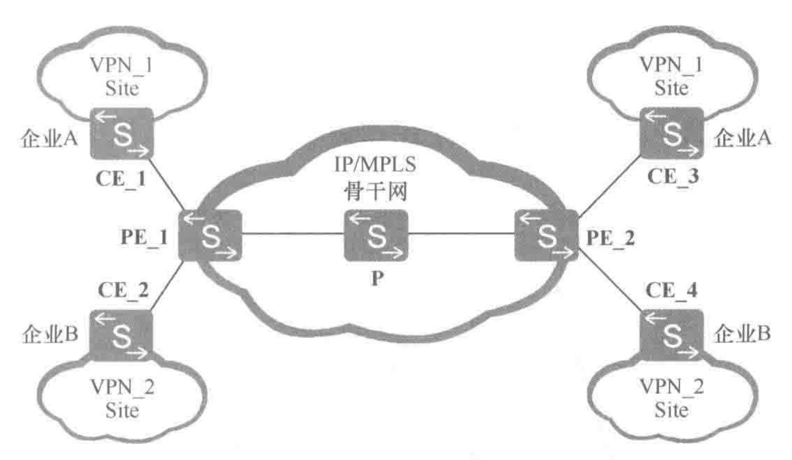


图 8-7　MPLS QoS 在区分不同 VPN 的优先级方面的应用示例

当不同企业 VPN 业务进入 MPLS 网络时，需要在 MPLS 网络中对不同企业 VPN 区分优先级，保证企业 A 的优先级高、企业 B 的优先级低，并根据优先级的高低对不同企业提供不同的 QoS 服务。

下面以图 8-7 中 L3VPN 报文从 PE_1 发往 PE_2 的过程为例介绍 MPLS QoS 在区不同 VPN 的优先级方面的应用。

（1）入节点（PE_1）将企业 A 和 B 的优先级分别标记为 MPLS 报文的 EXP 优先级，且企业 A 的优先级高于企业 B 的优先级，以便后续 MPLS 网络根据 EXP 优先级进行服务。

（2）中间节点（P）根据本地配置，将接收到的不同 VPN 中的 MPLS 报文携带的 EXP 优先级映射到不同的内部服务等级和颜色，再根据这些服务等级和颜色对报文进行不同的 QoS 服务。报文在流出设备时，设备根据不同 VPN 网络中的内部服务等级和颜色重标记为不同的 EXP 优先级。

（3）出节点（PE_2）根据本地配置，将接收到的不同 VPN 中的报文携带的 EXP 或 DSCP 优先级（具体要视 MPLS DiffServ 工作模式，以及倒数第二跳分配的 MPLS 标签是否支持 PHP 而定）映射到内部服务等级和颜色，再根这些据服务等级和颜色对报文进行不同的 QoS 服务，设备根据不同的内部服务等级和颜色重标记为不同的 IP 报文 DSCP 优先级，以便后续网络根据报文优先级进行服务。

8.2　MPLS QoS 配置与管理

S 系列交换机 MPLS QoS 功能的配置比较简单，主要存在以下两项配置任务。

1. 配置 MPLS 公网隧道标签优先级映射

这是基本 MPLS 公网隧道进行的 MPLS QoS 配置，通过配置不同 MPLS 公网隧道标

签和 EXP 的优先级映射关系，确定在不同 MPLS 公网隧道中传输的流量所具有不同优先级。

2. 配置 MPLS 私网支持的 DiffServ 模式

除了可以基于大范围的 MPLS 公网隧道优先级来进行业务优先级区分外，在 MPLS VPN 应用中，还可通过配置 VPN 所支持的 DiffServ 模式，实现在同一 MPLS 公网隧道相同或不同 VPN 网络内部不同业务的优先级区分。

8.2.1 配置 MPLS 公网隧道标签优先级映射

当需要对在同一 MPLS 公网隧道中传输的所有业务配置相同的 PHB 时，可配置 MPLS 公网隧道标签的优先级映射，不同的 MPLS 公网隧道可以配置不同的优先级映射关系，以实现在不同 MPLS 公网隧道中传输的流量具有不同的优先级。

说明 这里的 MPLS 公网隧道标签优先级映射配置还可应用于下节将要介绍的 MPLS 私网标签映射，因为私网 MPLS 标签优先级也可通过复制公网 MPLS 公网隧道标签中的优先级得到，如我们在 8.1.3 节中介绍的 Uniform、Pipe 和 Short pipe 三种模式均可在倒数第二跳时将弹出的外层 MPLS 标签中的 EXP 优先级复制到内层 MPLS 标签中，然后再根据其 EXP 优先级与对应的 PHB 进行映射。

另外，不管是骨干网中哪级节点，均可以选择采用 MPLS 公网隧道中配置的 DiffServ 域中的 EXP 与 PHB 映射关系进行映射。

MPLS 公网隧道标签优先级映射涉及两方面的配置任务：一是配置公网隧道所要使用的 DiffServ 域，并在其中配置 EXP 优先级与 PHB 之间的映射关系；二是在公网隧道上应用以上配置的 DiffServ 域，使通过公网隧道的所有业务流量都应用该 DiffServ 域中配置的 EXP 优先级与 PHB 的映射关系。

建议在配置 MPLS 公网隧道和 MPLS VPN 之前完成公网隧道标签优先级映射，否则配置好后还要重启信令协议，使配置生效。

1. 创建 DiffServ 域并配置优先级映射关系

DiffServ 域是由一组相连的 DiffServ 节点组成的，这些相连的 DiffServ 节点采用相同的服务提供策略并实现相同 PHB 组集合。配置好 DiffServ 域中 EXP 优先级与 PHB 的映射关系后，则使用该域的所有业务流量都将采用该域中的优先级映射配置。

要注意的是，在接口的不同方向上，所应用的 EXP 优先级与 PHB 映射关系是不一样的：当业务流流入设备（在接口入方向上）时，设备将报文携带的 EXP 优先级信息映射到相应的 PHB 行为、颜色，然后在设备内部再根据报文映射后的 PHB 行为进行拥塞管理，根据报文映射后的颜色进行拥塞避免；当业务流流出设备（在接口出方向上）时，设备再将报文映射后的 PHB 行为、颜色反向映射为相应的 EXP 优先级，对端设备根据报文的 EXP 优先级提供相应的 QoS 服务。

创建 DiffServ 域并配置 EXP 优先级与 PHB 的映射关系的具体配置步骤如表 8-2 所示，缺省的 EXP 到 PHB 行为颜色映射关系见表 8-3，缺省的 PHB 行为颜色到 EXP 优先级之间的映射关系见表 8-4。

表 8-2　　创建 **DiffServ** 域并配置 **EXP** 优先级与 **PHB** 的映射关系的步骤

步骤	命令	说明
1	**system-view** 例如：<HUAWEI> **system-view**	进入系统视图
2	**diffserv domain** { **default** | *ds-domain-name* } 例如：[HUAWEI] **diffserv domain** ds1	创建 DiffServ 域并进入 DiffServ 域视图。命令中的参数和选项说明如下。 • **default**：二选一选项，指定选用系统预先设定的缺省 DiffServ 域 • *ds-domain-name*：二选一参数，指定要创建的 DiffServ 域的名称，字符串形式，**区分大小写**，不支持空格，不能为“n”“no”“non”“none”，长度范围是 1～31。当输入的字符串两端使用双引号时，可在字符串中输入空格。DiffServ 域的名称不能为“--”。 缺省情况下，系统预定义了一个名为 default 的 DiffServ 域，default 域定义了缺省情况下报文的优先级和 PHB 行为、颜色之间的映射关系。用户可以修改 default 域中定义的映射关系，但不能删除 default 域。可用 **undo diffserv domain** *ds-domain-name* 命令删除指定的 DiffServ 域
3	**mpls-exp-inbound** *exp-value* **phb** *service-class* [*color*] 例如：[HUAWEI-dsdomain-ds1] **mpls-exp-inbound** 2 **phb af1 yellow**	（可选）在接口入方向，将 MPLS 报文的 EXP 优先级映射为 PHB 行为，并为报文着色。命令中的参数说明如下。 • *exp-value*：表示 MPLS 报文中携带的 EXP 优先级值，整数形式，取值范围是 0～7，值越大优先级越高。 • *service-class*：表示 PHB 行为，取值可以为 BE、AF1～AF4、EF、CS6 或 CS7，优先级依次提高。 • *color*：可选参数，表示报文标记的颜色，取值可以为 green、yellow 或 red，用于进行拥塞管理，不同拥塞管理方式对不同颜色报文的处理方式不一样，具体参见《华为交换机学习指南》一书的 QoS 拥塞管理部分。 缺省的 EXP 到 PHB 行为/颜色映射关系如表 8-3 所示，可用 **undo mpls-exp-inbound** 命令恢复缺省的映射关系，也可通过本命令修改映射关系
4	**mpls-exp-outbound** *service-class color* **map** *exp-value* 例如：[HUAWEI-dsdomain-ds1] **mpls-exp-outbound af1 yellow map** 2	（可选）在接口出方向，将 PHB 行为、颜色映射为 MPLS 报文的 EXP 优先级。命令中的参数说明参见以上命令中对应的参数。 缺省的 PHB 行为、颜色到 EXP 优先级之间的映射关系如表 8-4 所示，可用 **undo mpls-exp-outbound** 命令恢复缺省的映射关系，也可通过本命令修改映射关系

表 8-3　　缺省的 **EXP** 到 **PHB** 行为/颜色映射关系

EXP 优先级	PHB 行为	颜色
0	BE	green
1	AF1	green
2	AF2	green
3	AF3	green
4	AF4	green
5	EF	green
6	CS6	green
7	CS7	green

表 8-4 缺省的 **PHB** 行为、颜色到 **EXP** 优先级之间的映射关系

PHB 行为	颜色	EXP 优先级
BE	green	0
	yellow	0
	red	0
AF1	green	1
	yellow	1
	red	1
AF2	green	2
	yellow	2
	red	2
AF3	green	3
	yellow	3
	red	3
AF4	green	4
	yellow	4
	red	4
EF	green	5
	yellow	5
	red	5
CS6	green	6
	yellow	6
	red	6
CS7	green	7
	yellow	7
	red	7

2. 配置公网隧道优先级映射

当需要根据 DiffServ 域中定义的映射关系，对流入设备的报文进行优先级到 PHB 行为和颜色之间的映射操作时，可以将 DiffServ 域绑定到报文的入接口，系统会根据 DiffServ 域中的映射关系将报文的优先级映射为相应的 PHB 行为和颜色。

当需要根据 DiffServ 域中定义的映射关系，对流出设备的报文进行 PHB 行为到优先级之间的映射操作时，可以将 DiffServ 域绑定到报文的出接口，系统会根据 DiffServ 域中的映射关系将报文的 PHB 行为和颜色映射为优先级。

公网隧道优先级映射在不同节点上的配置命令有所不同，具体如表 8-5 所示。

表 8-5 配置公网隧道优先级映射的步骤

步骤	命令	说明
1	**system-view** 例如：<HUAWEI> **system-view**	进入系统视图
2	**mpls-qos ingress** { **use vpn-label-exp** | **trust upstream** { *ds-name* | **default** | **none** } } 例如：[HUAWEI] **mpls-qos ingress trust upstream** ds1	（多选一）在 Ingress 设备上配置公网隧道进行 EXP 的优先级映射。命令的参数和选项说明如下。 ● **use vpn-label-exp**：二选一选项，使用内层 EXP 标签值。如果希望根据私网隧道的 EXP 进行优先级映射，选择此选项。

（续表）

步骤	命令	说明
2	**mpls-qos ingress { use vpn-label-exp | trust upstream {** *ds-name* **| default | none } }** 例如：[HUAWEI] **mpls-qos ingress trust upstream** ds1	• **trust upstream**：二选项一选项，信任后面指定的 DiffServ 域。 • *ds-name*：多选一参数，指定所信任的 DiffServ 域名。 • **default**：多选一选项，指定信任的 DiffServ 域为 default 域。 • **none**：多选一选项，指定报文不进行公网隧道的 EXP 优先级映射，并将公网隧道的 EXP 设置为 0。 缺省情况下，根据缺省的 default 域进行公网隧道的 EXP 的优先级映射，可用 **undo mpls-qos ingress { use vpn-label-exp | trust upstream }**或 **undo mpls-qos ingress trust upstream none** 命令恢复缺省配置
	mpls-qos transit trust upstream { *ds-name* **| default | none }** 例如：[HUAWEI] **mpls-qos transit trust upstream** ds1	（多选一）在 Transit 设备上配置公网隧道基于 EXP 进行优先级映射。命令中的参数和选项说明参见以上 **mpls-qos ingress** 命令中的对应参数和选项。 缺省情况下，根据缺省的 default 域进行公网隧道的 EXP 的优先级映射，可用 **undo mpls-qos transit trust upstream** 或 **undo mpls-qos transit trust upstream none** 命令恢复缺省配置
	mpls-qos egress trust upstream { *ds-name* **| default | none }** 例如：[HUAWEI] **mpls-qos egress trust upstream** ds1	（多选一）在 Egress 设备上配置公网隧道基于 EXP 进行优先级映射。命令中的参数和选项说明参见以上 **mpls-qos ingress** 命令中的对应参数和选项。 缺省情况下，根据缺省的 default 域进行公网隧道的 EXP 的优先级映射，可用 **undo mpls-qos egress trust upstream** 或 **undo mpls-qos egress trust upstream none** 命令恢复缺省配置

以上配置完成后，可执行以下两 display 命令查看相关配置。

■ **display mpls l2vc** [*vc-id* | **interface** *interface-type interface-number* | **remote-info** [*vc-id* | **verbose**] | **state** { **down** | **Up** }]：查看 VLL 下 MPLS DiffServ 信息。

■ **display vsi** [**name** *vsi-name*] [**verbose**]：查看 VPLS 下 MPLS DiffServ 信息。

8.2.2 配置 MPLS 私网支持的 DiffServ 模式

上节介绍的 MPLS 公网隧道标签优先级映射可直接应用于在同一 MPLS 公网隧道中传输的报文，采用相同的 EXP 与 PHB 映射关系情形，同时也可应用于本节将要介绍的在 L2VPN 或者 L3VPN 网络中，不同 VPN，或者相同 VPN 中不同的业务流的 EXP 与 PHB 映射。但此时仍需要配置 MPLS 私网支持的 DiffServ 模式。

说明 有关 MPLS VPN，如 L3VPN 中的 BGP/MPLS IP VPN、L2VPN 中的 VLL（Virtual Leased Line，虚拟租用线路）、PWE3（Pseudo-Wire Emulation Edge to Edge 3，端到端伪线仿真 3）和 VPLS（Virtual Private LAN Service，虚拟专用局域网业务）等的具体技术原理和配置与管理方法请参见配套的《华为 MPLS VPN 学习指南》一书。

1. 配置 MPLS L2VPN 支持 DiffServ 模式

大部分 L2VPN 方案中可以在同一 MPLS 公网隧道中建立多条 PW，传输不同用户站点的私网二层报文。这时可根据不同的需求选择不同的 MPLS QoS 配置方式。

■ 如果用户希望在同一 VPN 内区分不同业务的优先级，可以配置差分服务模式为 Uniform；也可以配置差分服务模式为 Pipe 或 Short pipe，但此时必须指定引用的 Diffserv 域，以便在报文流出 PE 时采用对应 Diffserv 域中配置的 EXP 与 PHB 映射关系，而在 Uniform 模式中在报文流出 PE 时会直接通过报文中的 EXP 优先级改变报文中的 IP 网络优先级（可以是 DSCP、802.1p 优先级），决定优先级所映射的 PHB。

■ 如果用户不希望在同一 VPN 内区分不同业务的优先级，但是希望区分不同 VPN 的优先级时，可以配置差分服务模式为 Pipe 或 Short pipe，但此时必须指定私网标签的 EXP 优先级值，因为这两种模式中 MPLS 报文的 EXP 优先级是由用户指定的，并且不影响报文中的 IP 网络优先级，而 Uniform 模式中 MPLS 报文的 EXP 优先级是通过报文中的 IP 网络优先级映射得到。

说明 用户希望不改变原报文携带的优先级时，建议使用 Pipe 或 Short pipe 模式，因为 Uniform 模式可能会改变原报文携带的优先级。另外，Uniform 和 Pipe 模式在 Egress 上根据报文的 EXP 优先级选择 PHB，而在 Short pipe 模式中，L2VPN 场景根据报文的 802.1p 优先级，L3VPN 场景根据报文的 DSCP 优先级选择 PHB。

在 VLL 组网环境和 VPLS 组网模式中，MPLS 私网支持的 DiffServ 模式的配置方法总体差不多，只不过需要在不同视图下配置。在 VLL 组网模式中的配置方法如表 8-6 所示，在 VPLS 组网模式中的配置方法如表 8-7 所示。

表 8-6 **VLL 组网模式下的 DiffServ 模式配置步骤**

步骤	命令	说明
1	**system-view** 例如：<HUAWEI> **system-view**	进入系统视图
2	**interface** *interface-type interface-number* 例如：[HUAWEI] **interface vlanif** 100	进入 AC 接口的接口视图
3	**diffserv-mode** { **pipe** { **mpls-exp** *mpls-exp* | **domain** *ds-name* } | **short-pipe** [**mpls-exp** *mpls-exp*] **domain** *ds-name* | **uniform** [**domain** *ds-name*] } 例如：[HUAWEI-Vlanif100] **diffserv-mode pipe mpls-exp** 3	配置 VLL 支持的 Diffserv 模式。必须在 VC 建立之前配置该命令才会生效，否则需要对绑定的 AC 接口执行去绑定/绑定操作。命令中的参数和选项说明如下。 • **pipe**：多选一选项，指定 MPLS 的差分服务模式为 Pipe。 • **short-pipe**：多选一选项，指定 MPLS 的差分服务模式为 Short pipe。 • **uniform**：多选一选项，指定 MPLS 的差分服务模式为 uniform。 【说明】在 Ingress 对以上三种模式均会起作用，Egress 只会对 Short pipe 模式起作用。 • **mpls-exp** *mpls-exp*：二选一参数，指定私网标签的优先级映射值，值越大表示优先级越高，报文转发的质量越高。只会在 Ingress PE 的 Pipe、Short pipe 模式中起作用，在

（续表）

步骤	命令	说明
3	**diffserv-mode** { **pipe** { **mpls-exp** *mpls-exp* | **domain** *ds-name* } | **short-pipe** [**mpls-exp** *mpls-exp*] **domain** *ds-name* | **uniform** [**domain** *ds-name*] } 例如：[HUAWEI-Vlanif100] **diffserv-mode pipe mpls-exp** 3	Egress PE 不起作用；如果在 short-pipe 模式配置中同时选择了 domain *ds-name* 参数配置 Diffserv 域，则优先选取本参数值映射内层标签。 ● **domain** *ds-name*：指定引用的 Diffserv 域名，必须是已存在的 Diffserv 域名。缺省的 DiffServ 域名为 default。 【注意】在配置 VLL 支持的 Diffserv 模式中，需要注意以下事项。 ● 如果在 MPLS 公网隧道上配置了 **mpls-qos ingress trust upstream none** 或 **mpls-qos egress trust upstream none** 命令，即使配置了 **diffserv-mode** 命令，私网也不进行 EXP 优先级映射。 ● Ingress 节点上指定差分服务模式为 Uniform 时，如果不指定 **domain**，则根据公网 MPLS 隧道上配置的 **mpls-书 qos ingress trust upstream** { *ds-name* | **default** }命令指定的 Diffserv 域进行优先级映射；否则根据本命令指定的 Diffserv 域进行优先级映射。 ● Egress 节点上非 PHP 场景根据公网 MPLS 隧道上配置的 **mpls-qos egress trust upstream** { *ds-name* | **default** }命令指定的 Diffserv 域进行优先级映射；PHP 场景如果指定 **domain** 参数，则根据指定的 Diffserv 域进行优先级映射，否则根据公网 MPLS 隧道上配置的 **mpls-qos egress trust upstream** { *ds-name* | **default** }命令指定的 Diffserv 域进行优先级映射。 缺省情况下，MPLS L2VPN 私网标签的优先级映射服务模式是 Uniform 模式，可用 **undo diffserv-mode** 命令恢复缺省配置

表 8-7 VPLS 组网模式下的 DiffServ 模式配置步骤

步骤	命令	说明
1	**system-view** 例如：<HUAWEI> **system-view**	进入系统视图
2	**vsi** *vsi-name* 例如：[HUAWEI] **vsi** company1	进入指定的 VSI 实例视图
3	**diffserv-mode** { **pipe** { **mpls-exp** *mpls-exp* | **domain** *ds-name* } | **short-pipe** [**mpls-exp** *mpls-exp*] **domain** *ds-name* | **uniform** [**domain** *ds-name*] } 例如：[[HUAWEI-vsi-company1] **diffserv-mode pipe mpls-exp** 3	配置 VPLS 支持的 Diffserv 模式。必须在 VSI 实例生效之前配置该命令才会生效，否则需要对 VSI 执行禁止/使能操作。命令中的参数和选项说明，以及其他注意事项参见表 8-6 中第 3 步。 缺省情况下，MPLS L2VPN 私网标签的优先级映射服务模式是 Uniform 模式，可用 **undo diffserv-mode** 命令恢复缺省配置

2. 配置 MPLS L3VPN 支持 DiffServ 模式

在 L3VPN 中，也可以针对不同 VPN 网络，或者同一 VPN 网络中不同业务配置 MPLS QoS。在 MPLS DiffServ 工作模式选择方面也要区分“在同一 VPN 内区分不同业务的优

先级”和“在同一 VPN 内区分不同业务的优先级，但是希望区分不同 VPN 的优先级”两种情形，具体的选择原则与前面介绍的 L2VPN 场景下的介绍基本一样，参见即可，只是在 Short pipe 模式下要根据报文的 DSCP 优先级选择 PHB。

配置 MPLS L3VPN 支持 DiffServ 模式的步骤如表 8-8 所示。

表 8-8 **MPLS L3VPN 支持 DiffServ 模式的配置步骤**

步骤	命令	说明
1	**system-view** 例如：<HUAWEI> **system-view**	进入系统视图
2	**ip vpn-instance** *vpn-instance-name* 例如：[HUAWEI] **ip vpn-instance** vrf1	进入指定的 VPN 实例视图
3	**diffserv-mode** { **pipe** { **mpls-exp** *mpls-exp* | **domain** *ds-name* } | **short-pipe** [**mpls-exp** *mpls-exp*] **domain** *ds-name* | **uniform** [**domain** *ds-name*] } 例如：[HUAWEI-vpn-instance-vrf1] **diffserv-mode pipe mpls-exp** 3	配置 MPLS L3VPN 支持的 Diffserv 模式。必须在 VPN 实例生效之前配置该命令才会生效，否则需要复位 BGP 连接。命令中的参数和选项，以及注意事项均可参见表 8-6 中的第 3 步。 缺省情况下，MPLS L3VPN 支持 DiffServ 模式为 Uniform 模式，可用 **undo diffserv-mode** 命令恢复缺省配置

8.2.3 L3VPN MPLS QoS 配置示例

如图 8-8 所示，企业 A 和企业 B 通过部署 BGP/MPLS IP VPN，实现总部和分支机构的互联。CE1、CE3 连接企业 A 的总部和分支，CE2、CE4 连接企业 B 的总部和分支。企业 A 使用 vpna 实例，企业 B 使用 vpnb 实例。由于企业 A 的服务等级高，现要求给企业 A 提供更好的 QoS 保证。

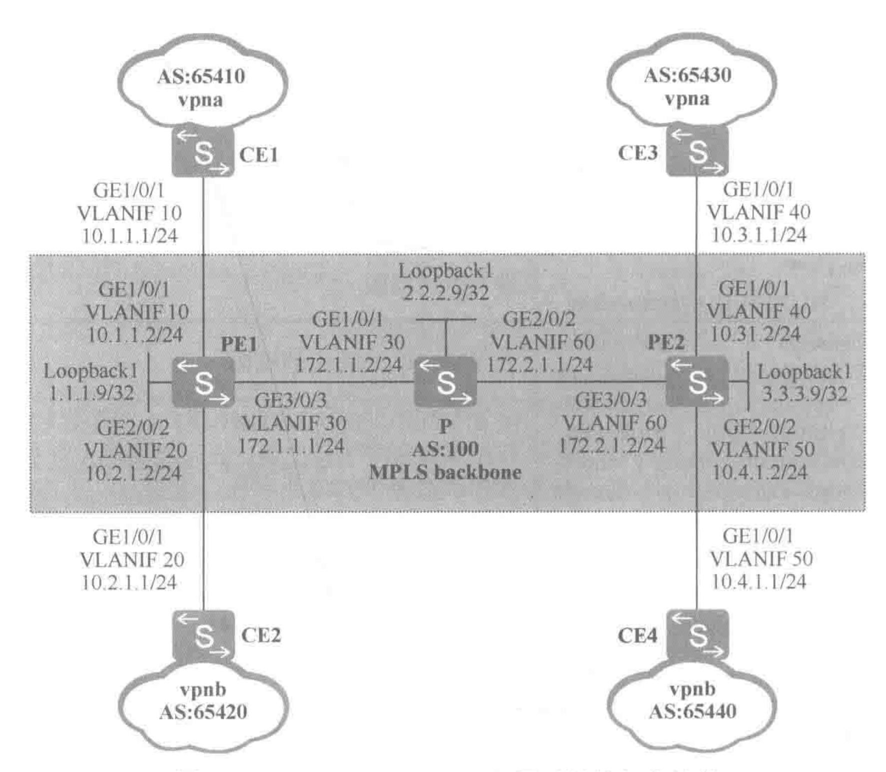


图 8-8 L3VPN MPLS QoS 配置示例的拓扑结构

1. 基本配置思路分析

MPLS QoS 配置其实简单来说，就是要在 MPLS 或 MPLS VPN 基础上进行 QoS 配置，所以本示例首先也要进行 BGP/MPLS IP VPN 建立的配置，然后再通过标签映射，对两个企业（对应不同的 VPN 实例）中的流量配置不同的 EXP 优先级，以提供不同的服务等级（企业 A 的要高于企业 B 的）。

本示例是 L3VPN 私网 VPN 场景，有关 BGP/MPLS IP VPN 的具体配置方法参见配套的《华为 MPLS VPN 学习指南》。本示例并没有说明两企业用户发送的报文中所带的 DSCP 优先值，所以如果采用 Uniform 模式时就无法确保企业 A 用户发送的报文具有更高优先级，故此时可以选择可以直接为这两家企业 VPN 实例配置不同的 MPLS 标签 EXP 优先级（可采用 Pipe 或 Short pipe 模式）。

在 Pipe 或 Short pipe 模式 DiffServ 中，Pipe 模式中报文进入到目的站点时的 PHB 行为、颜色可由报文中的 EXP 优先级决定，而在 Short pipe 模式中报文进入到目的站点时的 PHB 行为、颜色仍由原始 IP 报文中携带的 DSCP 优先级决定，本示例选择更佳的 Pipe 模式，使得报文中携带的 EXP 优先不仅影响报文在 MPLS 网络中的传输，还影响报文在离开 MPLS 网络后到达目的站点的途中的传输，确保企业 A 中的流量总是可以得到更高的服务等级。

本示例是 BGP/MPLS IP VPN，所以首先要按照《华为 MPLS VPN 学习指南》一书中的介绍配置好基本的 BGP/MPLS IP VPN 功能（PE 与 CE 间采用 EBGP 连接），然后再根据 8.2.2 节介绍的 L3VPN 场景下的 Pipe DiffServ 模式配置方法，在 PE1 和 PE2 上配置 MPLS QoS。具体配置思路如下。

（1）按图所示创建好各设备上所需的 VLAN，并把各接口加入到对应的 VLAN 中，配置各 VLANIF 接口和 Loopback 接口的 IP 地址和骨干网各节点的 OSPF 路由，实现骨干网三层互通。

（2）在骨干网各节点上配置 MPLS 基本能力和 LDP，建立公网 LSP 隧道。

（3）配置 PE1 与 PE2 之间的 MP-IBGP 会话，为两企业中的私网路由分配私网标签。也使两 PE 间能直接交互 Update 消息相互，通告彼此所连接的内部私网 VPN-IPv4 路由。

（4）在两 PE 上创建两企业站点对应的 VPN 实例，并绑定在对应的 PE 连接 CE 的 AC 接口上，配置其他相关 VPN 属性。同一 VPN 中各 VPN 实例的 VPN-Target 属性中，至少要使入方向的 VPN-Target 与对端 PE 上配置出方向 VPN-Target 属性要匹配，RD 属性值配置要各不一样。

（5）在 PE 与 CE 之间建立 EBGP 对等体关系，在 PE 上引入 VPN 私网路由（本示例只需在 BGP 路由进程中引入 PE 与 CE 直连的路由即可）。

（6）在 Ingress 节点上配置 MPLS QoS。在双向通信中，PE1 和 PE2 均可能为 Ingress，所以需要分别配置。其中，vpna 和 vpnb 实例均采用 Pipe 模式，采用缺省的 default Diffserv 域中的 EXP 和 PHB 映射关系，两 VPN 实例的 EXP 优先级值分别设置为 4 和 3，以实现对企业 A 的业务提供更好的 QoS 保证。

说明 以上6个步骤中，第（1）~（2）步是用于完成基本的公网 MPLS LSP 隧道建立；第（3）~（5）步是基本的 BGP/MPLS IP VPN 配置；第（6）步才是最终的 MPLS QoS 配置。

2. 具体配置步骤

（1）在各设备上创建所需 VLAN，并把各接口加入到对应的 VLAN 中，同时配置 MPLS 骨干网上各节点的 OSPF 协议，实现骨干网三层互通。

\# PE1 上的配置。

```
<HUAWEI> system-view
[HUAWEI] sysname PE1
[PE1] interface loopback 1
[PE1-LoopBack1] ip address 1.1.1.9 32
[PE1-LoopBack1] quit
[PE1] vlan batch 10 20 30
[PE1] interface gigabitethernet 1/0/1
[PE1-GigabitEthernet1/0/1] port link-type trunk
[PE1-GigabitEthernet1/0/1] port trunk allow-pass vlan 10
[PE1-GigabitEthernet1/0/1] quit
[PE1] interface gigabitethernet 2/0/2
[PE1-GigabitEthernet2/0/2] port link-type trunk
[PE1-GigabitEthernet2/0/2] port trunk allow-pass vlan 20
[PE1-GigabitEthernet2/0/2] quit
[PE1] interface gigabitethernet 3/0/3
[PE1-GigabitEthernet3/0/3] port link-type trunk
[PE1-GigabitEthernet3/0/3] port trunk allow-pass vlan 30
[PE1-GigabitEthernet3/0/3] quit
[PE1] interface vlanif 30
[PE1-Vlanif30] ip address 172.1.1.1 24
[PE1-Vlanif30] quit
[PE1] ospf 1
[PE1-ospf-1] area 0
[PE1-ospf-1-area-0.0.0.0] network 172.1.1.0 0.0.0.255
[PE1-ospf-1-area-0.0.0.0] network 1.1.1.9 0.0.0.0
[PE1-ospf-1-area-0.0.0.0] quit
[PE1-ospf-1] quit
```

\# P 上的配置。

```
<HUAWEI> system-view
[HUAWEI] sysname P
[P] interface loopback 1
[P-LoopBack1] ip address 2.2.2.9 32
[P-LoopBack1] quit
[P] vlan batch 30 60
[P] interface gigabitethernet 1/0/1
[P-GigabitEthernet1/0/1] port link-type trunk
[P-GigabitEthernet1/0/1] port trunk allow-pass vlan 30
[P-GigabitEthernet1/0/1] quit
[P] interface gigabitethernet 2/0/2
[P-GigabitEthernet2/0/2] port link-type trunk
[P-GigabitEthernet2/0/2] port trunk allow-pass vlan 60
[P-GigabitEthernet2/0/2] quit
[P] interface vlanif 30
```

```
[P-Vlanif30] ip address 172.1.1.2 24
[P-Vlanif30] quit
[P] interface vlanif 60
[P-Vlanif60] ip address 172.2.1.1 24
[P-Vlanif60] quit
[P] ospf
[P-ospf-1] area 0
[P-ospf-1-area-0.0.0.0] network 172.1.1.0 0.0.0.255
[P-ospf-1-area-0.0.0.0] network 172.2.1.0 0.0.0.255
[P-ospf-1-area-0.0.0.0] network 2.2.2.9 0.0.0.0
[P-ospf-1-area-0.0.0.0] quit
[P-ospf-1] quit
```

PE2 上的配置。

```
<HUAWEI> system-view
[HUAWEI] sysname PE2
[PE2] interface loopback 1
[PE2-LoopBack1] ip address 3.3.3.9 32
[PE2-LoopBack1] quit
[PE2] vlan batch 40 50 60
[PE2] interface gigabitethernet 1/0/1
[PE2-GigabitEthernet1/0/1] port link-type trunk
[PE2-GigabitEthernet1/0/1] port trunk allow-pass vlan 40
[PE2-GigabitEthernet1/0/1] quit
[PE2] interface gigabitethernet 2/0/2
[PE2-GigabitEthernet2/0/2] port link-type trunk
[PE2-GigabitEthernet2/0/2] port trunk allow-pass vlan 50
[PE2-GigabitEthernet2/0/2] quit
[PE2] interface gigabitethernet 3/0/3
[PE2-GigabitEthernet3/0/3] port link-type trunk
[PE2-GigabitEthernet3/0/3] port trunk allow-pass vlan 60
[PE2-GigabitEthernet3/0/3] quit
[PE2] interface vlanif 60
[PE2-Vlanif60] ip address 172.2.1.2 24
[PE2-Vlanif60] quit
[PE2] ospf 1
[PE2-ospf-1] area 0
[PE2-ospf-1-area-0.0.0.0] network 172.2.1.0 0.0.0.255
[PE2-ospf-1-area-0.0.0.0] network 3.3.3.9 0.0.0.0
[PE2-ospf-1-area-0.0.0.0] quit
[PE2-ospf-1] quit
```

CE1 上的配置。

```
<HUAWEI> system-view
[HUAWEI] sysname CE1
[CE1] vlan batch 10
[CE1] interface vlanif 10
[CE1-Vlanif10] ip address 10.1.1.1 255.255.255.0
[CE1-Vlanif10] quit
[CE1] interface GigabitEthernet1/0/1
[CE1-GigabitEthernet1/0/1] port link-type trunk
[CE1-GigabitEthernet1/0/1] port trunk allow-pass vlan 10
[CE1-GigabitEthernet1/0/1] quit
```

CE2 上的配置。

```
<HUAWEI> system-view
[HUAWEI] sysname CE2
```

```
[CE2] vlan batch 20
[CE2]interface vlanif 20
[CE2-Vlanif20] ip address 10.2.1.1 255.255.255.0
[CE2-Vlanif20] quit
[CE2] interface GigabitEthernet1/0/1
[CE2-GigabitEthernet1/0/1] port link-type trunk
[CE2-GigabitEthernet1/0/1] port trunk allow-pass vlan 20
[CE2-GigabitEthernet1/0/1] quit
```

CE3 上的配置。

```
<HUAWEI> system-view
[HUAWEI] sysname CE3
[CE3] vlan batch 40
[CE3] interface vlanif 40
[CE3-Vlanif40] ip address 10.3.1.1 255.255.255.0
[CE3-Vlanif40] quit
[CE3] interface GigabitEthernet1/0/1
[CE3-GigabitEthernet1/0/1] port link-type trunk
[CE3-GigabitEthernet1/0/1] port trunk allow-pass vlan 40
[CE3-GigabitEthernet1/0/1] quit
```

CE4 上的配置。

```
<HUAWEI> system-view
[HUAWEI] sysname CE4
[CE4] vlan batch 50
[CE4] interface vlanif 50
[CE4-Vlanif50] ip address 10.4.1.1 255.255.255.0
[CE4-Vlanif50] quit
[CE4] interface GigabitEthernet1/0/1
[CE4-GigabitEthernet1/0/1] port link-type trunk
[CE4-GigabitEthernet1/0/1] port trunk allow-pass vlan 50
[CE4-GigabitEthernet1/0/1] quit
```

以上配置完成后，PE1、P、PE2 之间应能建立 OSPF 邻居关系，执行 **display ip routing-table** 命令可以看到 PE 之间学习到对方的 Loopback1 路由。

（2）在 MPLS 骨干网上配置 MPLS 基本能力和 MPLS LDP，建立 LDP LSP 隧道。

PE1 上的配置。

```
[PE1] mpls lsr-id 1.1.1.9
[PE1] mpls
[PE1-mpls] quit
[PE1] mpls ldp
[PE1-mpls-ldp] quit
[PE1] interface vlanif 30
[PE1-Vlanif30] mpls
[PE1-Vlanif30] mpls ldp
[PE1-Vlanif30] quit
```

P 上的配置。

```
[P] mpls lsr-id 2.2.2.9
[P] mpls
[P-mpls] quit
[P] mpls ldp
[P-mpls-ldp] quit
[P] interface vlanif 30
[P-Vlanif30] mpls
[P-Vlanif30] mpls ldp
```

```
[P-Vlanif30] quit
[P] interface vlanif 60
[P-Vlanif60] mpls
[P-Vlanif60] mpls ldp
[P-Vlanif60] quit
```

#　PE2 上的配置。

```
[PE2] mpls lsr-id 3.3.3.9
[PE2] mpls
[PE2-mpls] quit
[PE2] mpls ldp
[PE2-mpls-ldp] quit
[PE2] interface vlanif 60
[PE2-Vlanif60] mpls
[PE2-Vlanif60] mpls ldp
[PE2-Vlanif60] quit
```

上述配置完成后，PE1 与 P、P 与 PE2 之间应能建立 LDP 会话，执行 **display mpls ldp session** 命令可以看到显示结果中 Status 项为 **"Operational"**。以下是在 PE1 上执行该命令的输出示例，参见输出信息中的粗体字部分。

```
[PE1] display mpls ldp session

 LDP Session(s) in Public Network
 Codes: LAM(Label Advertisement Mode), SsnAge Unit(DDDD:HH:MM)
 A '*' before a session means the session is being deleted.
 ------------------------------------------------------------------------------
 PeerID              Status       LAM  SsnRole  SsnAge        KASent/Rcv
 ------------------------------------------------------------------------------
 2.2.2.9:0           Operational DU    Active    0000:00:01   6/6
 ------------------------------------------------------------------------------
 TOTAL: 1 session(s) Found.
```

（3）在 PE 之间建立 MP-IBGP 对等体关系，使两 PE 间可直接交互 BGP Update 消息，分配私网路由标签。因为 PE1 和 PE2 是直连的，故所指定的连接源接口不是物理接口，而是可以代表设备本身的 Loopback 接口，对等体 IP 地址也即对端该 Loopback 接口的 IP 地址。

#　PE1 上的配置。

```
[PE1] bgp 100
[PE1-bgp] peer 3.3.3.9 as-number 100
[PE1-bgp] peer 3.3.3.9 connect-interface loopback 1
[PE1-bgp] ipv4-family vpnv4
[PE1-bgp-af-vpnv4] peer 3.3.3.9 enable
[PE1-bgp-af-vpnv4] quit
[PE1-bgp] quit
```

#　PE2 上的配置。

```
[PE2] bgp 100
[PE2-bgp] peer 1.1.1.9 as-number 100
[PE2-bgp] peer 1.1.1.9 connect-interface loopback 1
[PE2-bgp] ipv4-family vpnv4
[PE2-bgp-af-vpnv4] peer 1.1.1.9 enable
[PE2-bgp-af-vpnv4] quit
[PE2-bgp] quit
```

以上配置完成后，在 PE 设备上执行 **display bgp peer** 命令，可以看到 PE 之间的

BGP 对等体关系已建立，并达到 Established 状态。以下是在 PE1 上执行该命令的输出示例，参见输出信息中的粗体字部分。

```
[PE1] display bgp peer

 BGP local router ID : 1.1.1.9
 Local AS number : 100
 Total number of peers : 1                 Peers in established state : 1

  Peer       V     AS   MsgRcvd   MsgSent   OutQ   Up/Down        State        PrefRcv

  3.3.3.9          4    100       12        6      0 00:02:21     Established   0
```

（4）在两 PE 上为两企业创建 VPN 实例，并绑定在 PE 连接 CE 的对应担当 AC 接口的 VLANIF 接口上。

此处，假设两 PE 上创建的两个 VPN 实例的 VPN-Target 两方向属性值分别为 111:1、222:2，各 CE 站点的 RD 属性分别为 100:1、100:2、200:1、100:2，然后在对应的 AC 接口（本示例采用 VLANIF 接口作为 AC 接口）上绑定对应的 VPN 实例。

\# PE1 上的配置。

```
[PE1] ip vpn-instance vpna
[PE1-vpn-instance-vpna] ipv4-family
[PE1-vpn-instance-vpna-af-ipv4] route-distinguisher 100:1
[PE1-vpn-instance-vpna-af-ipv4] vpn-target 111:1 both
[PE1-vpn-instance-vpna-af-ipv4] quit
[PE1-vpn-instance-vpna] quit
[PE1] ip vpn-instance vpnb
[PE1-vpn-instance-vpnb] ipv4-family
[PE1-vpn-instance-vpnb-af-ipv4] route-distinguisher 100:2
[PE1-vpn-instance-vpnb-af-ipv4] vpn-target 222:2 both
[PE1-vpn-instance-vpnb-af-ipv4] quit
[PE1-vpn-instance-vpnb] quit
[PE1] interface vlanif 10
[PE1-Vlanif10] ip binding vpn-instance vpna
[PE1-Vlanif10] ip address 10.1.1.2 24
[PE1-Vlanif10] quit
[PE1] interface vlanif 20
[PE1-Vlanif20] ip binding vpn-instance vpnb
[PE1-Vlanif20] ip address 10.2.1.2 24
[PE1-Vlanif20] quit
```

\# PE2 上的配置。

```
[PE2] ip vpn-instance vpna
[PE2-vpn-instance-vpna] ipv4-family
[PE2-vpn-instance-vpna-af-ipv4] route-distinguisher 200:1
[PE2-vpn-instance-vpna-af-ipv4] vpn-target 111:1 both
[PE2-vpn-instance-vpna-af-ipv4] quit
[PE2-vpn-instance-vpna] quit
[PE2] ip vpn-instance vpnb
[PE2-vpn-instance-vpnb] ipv4-family
[PE2-vpn-instance-vpnb-af-ipv4] route-distinguisher 200:2
[PE2-vpn-instance-vpnb-af-ipv4] vpn-target 222:2 both
[PE2-vpn-instance-vpnb-af-ipv4] quit
[PE2-vpn-instance-vpnb] quit
[PE2] interface vlanif 40
```

```
[PE2-Vlanif40] ip binding vpn-instance vpna
[PE2-Vlanif40] ip address 10.3.1.2 24
[PE2-Vlanif40] quit
[PE2] interface vlanif 50
[PE2-Vlanif50] ip binding vpn-instance vpnb
[PE2-Vlanif50] ip address 10.4.1.2 24
[PE2-Vlanif50] quit
```

（5）在 PE 与 CE 之间建立 EBGP 对等体关系，引入 VPN 私网路由。

CE1 上的配置。

```
[CE1] bgp 65410
[CE1-bgp] peer 10.1.1.2 as-number 100
[CE1-bgp] import-route direct    #---引入直连路由
```

CE2 上的配置。

```
[CE2] bgp 65420
[CE2-bgp] peer 10.2.1.2 as-number 100
[CE2-bgp] import-route direct
```

CE3 上的配置。

```
[CE3] bgp 65430
[CE3-bgp] peer 10.3.1.2 as-number 100
[CE3-bgp] import-route direct
```

CE4 上的配置。

```
[CE4] bgp 65440
[CE4-bgp] peer 10.4.1.2 as-number 100
[CE4-bgp] import-route direct
```

PE1 上的配置。

```
[PE1] bgp 100
[PE1-bgp] ipv4-family vpn-instance vpna
[PE1-bgp-vpna] peer 10.1.1.1 as-number 65410
[PE1-bgp-vpna] import-route direct
[PE1-bgp-vpna] quit
[PE1-bgp] ipv4-family vpn-instance vpnb
[PE1-bgp-vpnb] peer 10.2.1.1 as-number 65420
[PE1-bgp-vpnb] import-route direct
[PE1-bgp-vpnb] quit
[PE1-bgp] quit
```

PE2 上的配置。

```
[PE2] bgp 100
[PE2-bgp] ipv4-family vpn-instance vpna
[PE2-bgp-vpna] peer 10.3.1.1 as-number 65430
[PE2-bgp-vpna] import-route direct
[PE2-bgp-vpna] quit
[PE2-bgp] ipv4-family vpn-instance vpnb
[PE2-bgp-vpnb] peer 10.4.1.1 as-number 65440
[PE2-bgp-vpnb] import-route direct
[PE2-bgp-vpnb] quit
[PE2-bgp] quit
```

以上配置完成后，在 PE 设备上执行 **display bgp vpnv4 vpn-instance peer** 命令，可以看到 PE 与 CE 之间的 BGP 对等体关系已建立，并达到 Established 状态。以下是在 PE1 上执行该命令的输出示例。

```
[PE1] display bgp vpnv4 vpn-instance vpna peer
```

```
BGP local router ID : 1.1.1.9
Local AS number : 100
Total number of peers : 1                Peers in established state : 1

 Peer          V    AS   MsgRcvd  MsgSent  OutQ  Up/Down     State        PrefRcv

 10.1.1.1      4 65410       11        9     0 00:07:25   Established        1
```

（6）配置 MPLS QoS，采用 Pipe DiffServ 模式，在公网隧道进行 EXP 优先级映射时选择内层 VPN 标签中的 EXP 优先级，采用缺省的 default Diffserv 域，并为 vpna 和 vpnb 实例分别配置 EXP 优先级值为 4、3（缺省所映射的 PHB 行为/颜色分别为 AF4/green、AF3/green），使企业 A 的业务优先级高于企业 B 的。

PE1 上的配置。

```
[PE1] mpls-qos ingress use vpn-label-exp    #---指定内层 MPLS 标签中的 EXP 优先级值与 PHB 进行映射
[PE1] ip vpn-instance vpna
[PE1-vpn-instance-vpna] diffserv-mode pipe mpls-exp 4   #---设置 vpna 实例中的内层 MPLS 标签的 EXP 优先级为 4，采用 pipe 模式
[PE1-vpn-instance-vpna] quit
[PE1] ip vpn-instance vpnb
[PE1-vpn-instance-vpnb] diffserv-mode pipe mpls-exp 3    #---设置 vpnb 实例中的内层 MPLS 标签的 EXP 优先级为 3，采用 pipe 模式
[PE1-vpn-instance-vpnb] quit
```

PE2 上的配置。

```
[PE2] mpls-qos ingress use vpn-label-exp
[PE2] ip vpn-instance vpna
[PE2-vpn-instance-vpna] diffserv-mode pipe mpls-exp 4
[PE2-vpn-instance-vpna] quit
[PE2] ip vpn-instance vpnb
[PE2-vpn-instance-vpnb] diffserv-mode pipe mpls-exp 3
[PE2-vpn-instance-vpnb] quit
```

以上配置完成后，重启 MPLS LDP 和复位 BGP 连接，使配置生效。



第9章 MPLS DS-TE配置与管理



第 8 章介绍了华为 S 系列交换机中的 MPLS QoS 功能技术原理及相关的功能配置与管理方法，本章介绍华为 AR G3 系列路由器中的 MPLS QoS 配置与管理方法。

在 AR G3 系列路由器中，MPLS TE 隧道提供了更强大的 QoS 功能，它是以 MPLS DS-TE 隧道技术来实现的。在第 5 章就已说到，MPLS TE 隧道是具有部分 MPLS QoS 功能的 MPLS 隧道，而 MPLS DS-TE 隧道是由 MPLS TE 隧道扩展而来的，具有更全面、更强大的 MPLS QoS 功能。

MPLS DS-TE 隧道与 MPLS TE 隧道相比，在 CR-LSP 建立方面具有更灵活、更严格的约束条件。如一条 TE 隧道最多只能提供 2 种业务类型区分，一条 DS-TE 隧道上最多可以进行 8 种业务类型区分，结合 MPLS EXP 优先级，最多可提供 64 种流量分类。DS-TE 隧道还可以为不同 CR-LSP 配置不同 LSP 抢占优先级、DS-TE 带宽约束模型和业务调度模式等。本章将对以上技术原理，以及相关功能的配置与管理进行全面介绍。

9.1 DS-TE 基础及工作原理

在第 5 章介绍的传统 MPLS TE 隧道主要是仅要求沿途的每个节点预留资源以保证服务质量，最多可为两类业务（BC0 和 BC1）配置可预留的最大带宽，无法为更多类型业务分别配置预留带宽，也无法为不同业务类型提供不同的优先级服务。

例如当同一条隧道同时承载有语音流和数据流时，由于语音流对时延更为敏感，因此会要求数据流比语音流具有更高的丢弃优先级。但传统的 MPLS TE 对语音流和数据流会分配相同的丢弃优先级，无法提供有区别的 QoS 保证。为了解决这一问题，就诞生了一种称之为 MPLS DS-TE（DiffServ-Aware Traffic Engineering，差分服务感知流量工程）的解决方案。它不仅可以在同一 TE 隧道上承载不同类型的业务，而且可为不同类型业务预留不同的资源，配置不同的优先级，这样就可保证高优先级的业务总是拥有更高的服务级别。

9.1.1 MPLS DS-TE 的产生背景

MPLS DS-TE 同时结合了 MPLS TE 和 MPLS DiffServ 这两项技术。MPLS DiffServ 已在第 8 章 8.1 节介绍，是差分服务模型 Diff-Serv（Differentiated Services）的扩展。IP 网络中的 Diff-Serv 模型可以根据业务的不同服务等级，有差别地进行流量的控制和转发。但是 Diff-Serv 模型只能在单个节点上预留资源，即在配置了 Diff-Serv 服务型的本地设备上为某类业务预留带宽资源，却无法在整个路径上保证服务质量。因为 Diff-Serv 服务模型是无信令信息传播的，除非在整条路径的设备上都配置这样的 Diff-Serv 服务，但这样配置的工作量非常大。所以，即使同时使用了 Diff-Serv 和 MPLS TE，也不能全部满足要求，需要对 Diff-Serv 进行扩展，得到了 MPLS DiffServ。

例如，在某个应用场景中，一条路径同时承载语音业务和数据业务。为保证语音业务的质量，需要降低语音流的总延迟，限制每一条链路上语音流量不超过一定的比例。如果对该场景同时使用传统的 Diff-Serv 和 MPLS TE，使用 Diff-Serv 模型区分出业务的类型，**再对每类业务使用单独的 MPLS TE 隧道承载**，则当网络发生链路/节点故障、网络拓扑变化或者 LSP 抢占时，链路上的语音流量仍可能超过一个最佳性能所需的比例（如

占满了整个链路总带宽），使语音业务的延迟得不到保证。

如图 9-1 所示，假设所有链路的带宽均为 100Mbit/s，且链路开销相同，且 Router_1→Router_3→Router_4，Router_2→Router_3→Router_7→Router_4 这两条路径上都有语音流，分别为 60Mbit/s 和 40Mbit/s。

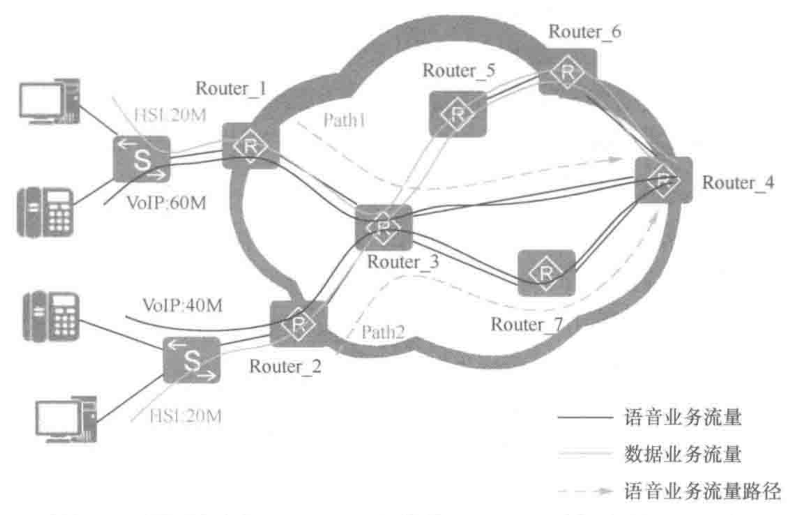


图 9-1 同时部署 MPLS TE 和传统 Diff-Serv 模型的组网示意

其中 Router_1→Router_3→Router_4 的语音流通过 Path1 的 MPLS TE 隧道传输，此时 Router_3→Router_4 链路上的语音流带宽百分比为 60%；Router_2→Router_3→Router_7→Router_4 的语音流通过 Path2 的 MPLS TE 隧道传输，此时 Router_3→Router_7→Router_4 上的语音流带宽百分比为 40%。而 Router_1 和 Router_2 发送的普通数据流均走 Rouer_3→Router_5→Router_6→Router_4 这条不同的路径。

当 Router_3 和 Router_4 之间的链路发生故障时，此时会根据 IGP 重新选择路径，最终原来的 Router_1→Router_3→Router_4 路径也改为经过 Router_7 的路径——Path3，因为此路径是带宽足够的最短路径，如图 9-2 所示。

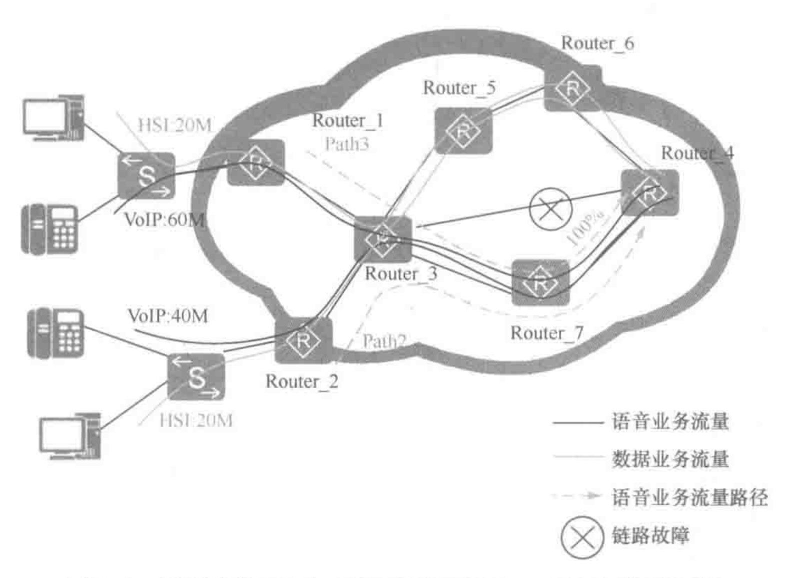


图 9-2 链路故障后 MPLS TE 和 DiffServ 方案的流量切换

此时，两路语音流都通过了 Router_3→Router_7→Router_4 路径，占满了这两段链路 100%的总带宽，导致语音流的总延时过长。但如果采用了 MPLS TE 和 MPLS DiffServ 的结合方案就可以解决以上问题了，因为扩展后的 MPLS DiffServ 可通过 RSVP-TE 信令协议在整条路径为各类业务流进行带宽预留，当某段链路的可用带宽小于所需的预留带宽时，会重新选择其他路径来重建 CR-LSP，而不是直接根据 IGP 路由选择最短的路径，这样可充分为各类业务流在整条路径上提供充分的带宽资源保障。

为了使 MPLS TE 能基于流量类型分配资源，提供差分服务，MPLS DS-TE 引入了 CT（Class Type）的概念。MPLS DS-TE 是解决骨干网 QoS 的有效技术，因为它可将 CR-LSP 的总带宽划分为 1～8 个部分，每部分被赋予不同的服务等级（0~7），允许基于 CoS（服务分类）粒度的资源预留，并提供在每个 CoS 级别的容错。一条 MPLS DS-TE 隧道中的一条或一组 CR-LSP 的相同服务等级的带宽集合称为一个 CT。

对于图 9-1 所示的场景，可以使用多 CT 的 CR-LSP，即将一条 CR-LSP 划分了多个 CT，同时承载不同服务等级的流量。此时，可以配置 Router_1→Router_3→Router_4 路径的 VoIP（语音）和 HIS（数据）流量使用同一条 MPLS TE 隧道的不同 CT 承载；Router_2→Router_3→Router_7→Router_4 的 VoIP 和 HIS 流量又一起使用另一条 MPLS TE 隧道的不同 CT 承载，这样就可以维持各类流量的相对比例，使各链路上语音流的带宽百分比控制在合理范围内（两个 CT 中的流量之和小于链路带宽 100Mbit/s），如图 9-3 所示。

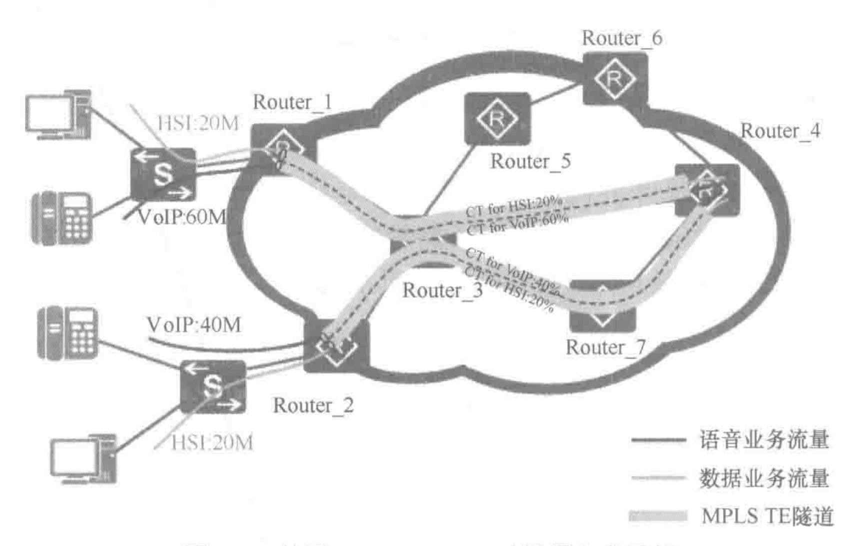


图 9-3 使用 MPLS DS-TE 时部署方案示例

当 Router_3→Router_4 之间链路出现了故障时，则 Router_1→Router_3→Router_4 这条路径下原来的 VoIP 和 HIS 流量会同时切换到 Router_1→Router_3→Router_5→Router_6→Router_4 这条路径，而不会再去挤 Router_3→Router_7→Router_4 这条路径了，否则两条 MPLS TE 隧道中的流量总和就大于链路总带宽 100Mbit/s，使得各类业务的传输受到影响。这样一来，切换之后 Router_1 到 Router_4 上的语音流带宽百分比依然可控制在合理的链路带宽 60%范围内。

9.1.2 MPLS DS-TE 基本概念

本节介绍一些与 MPLS DS-TE 相关的基本概念，以便更好地理解后面将要介绍的 MPLS DS-TE 技术原理。

（1）DS 字段

为了实现 Diff-Serv，RFC2474 中对 IPv4 报文头的 ToS 字段进行了重新定义，称为 DS（Differentiated Services，差分服务）字段。DS 字段的高 2 位是预留位，低 6 位是 DSCP（DS CodePoint，差分服务代码点）。有关 IPv4 的具体报头格式参见笔者《深入理解计算机网络》（新版）。

（2）PHB

PHB（Per Hop Behavior，每跳行为）用来描述拥有相同 DSCP 值的报文的下一步转发动作。目前，IETF 定义的三种标准 PHB 分别是 EF（Expedited Forwarding，加速转发）、AF（Assured Forwarding，确保转发）和 BE（Best-Effort，尽力而为），BE 是缺省的 PHB。而在每种 PHB 中又可分为三种不同的服务等级，具体参见《华为交换机学习指南》。

（3）CT

为了提供差分服务，DS-TE 将隧道中的一条或一组 CR-LSP 的带宽最多划分为 8 个部分，每部分被赋予不同的服务等级（0~7），相同服务等级的带宽集合称为一个 CT。一个 CT 只能承载具有相同服务等级的一种业务类型的流量，但可以包括一条 DS-TE 隧道中的一条或多条 CR-LSP。

IETF 规定一条 DS-TE 隧道中最多支持 8 个 CT，可以记为 CT*i*，其中 *i* 的取值范围是 0～7，对应其 CT 的服务等级（**数值越大，等级越高**）。这样一来，在一条 DS-TE 隧道中最多只支持 8 种不同类型的业务流。CT0 是最低服务级别的 CT，是为“尽力而为”类型业务提供的。

（4）IGP 的扩展

为支持 DS-TE，RFC4124 对 IGP 中的 OSPF 和 IS-IS 进行了扩展，形成了 OSPF-TE 和 ISIS-TE，分别参见第 5 章 5.2.2 节和 5.2.3 节。在这两种扩展 IGP 中，引入了新的可选子 TLV——带宽约束子 TLV（Bandwidth Constraints sub-TLV），并重新定义了原有的非预留带宽子 TLV（Unreserved bandwidth sub-TLV）的含义，用于通告和收集链路上各优先级的每个 CT 的可预留带宽。

（5）单 CT LSP 和多 CT LSP

单 CT LSP 是指一条 CR-LSP 中只有一个 CT，即只允许承载一种业务类型的流量，如静态 CR-LSP 就是这种类型。

多 CT LSP 是指一条 CR-LSP 中划分了多个 CT，可同时承载多种不同业务类型的流量。但在多 CT LSP 中，必须所有 CT 带宽都满足时，资源预留、CR-LSP 建立或带宽抢占才能成功。

9.1.3 LSP 抢占和 TE-Class 映射

与第 5 章介绍的 MPLS-TE 一样，DS-TE 中在建立 CR-LSP 的过程中，如果无法找到满足所需带宽要求的路径，则拆除另外一条已经建立的路径，抢占为它分配的带宽资

源。DS-TE 也是使用 MPLS TE 中的两个优先级属性来决定是否可以进行抢占：建立优先级（setup-priority）和保持优先级（hold-priority），可统称为抢占优先级。

TE-Class（TE 分类）是 CT 与抢占优先级的组合<CT，优先级>，用于统一管理 CT 与抢占优先级之间的映射。不同 CR-LSP 中的相同 CT 可以具有相同或不同的抢占优先级。TE-Class 与 CT 之间的关系可以描述为 TE-Class[*n*]=<CT*i*，*priority*>。

- *n* 用来标识 TE-Class，不与 CT 中 *i* 一一对应。
- *i* 用来标识 CT 类型，也即 CT 的服务等级；*priority* 为分配给对应类型 CT 的抢占优先级，可以是建立优先级，或保持优先级，但同一 CT 的建立优先级不能高于其保持优先级。
- $0 \leqslant i \leqslant 7$，$0 \leqslant priority \leqslant 7$。

抢占优先级的取值范围是 0～7，**数值越小，优先级越高**（这与 EXP 优先级相反）。只有当一条 CR-LSP 的 CT 和建立优先级的组合<CT，*setup-priority*>，以及 CT 和保持优先级的组合<CT，*hold-priority*>同时存在于 TE-Class（流量工程类）映射表中，该 CR-LSP 才能建立成功。如某节点的 TE-Class 映射表中仅有 TE-Class[0]＝<CT0，6>和 TE-Class[1]＝<CT0，7>，则只有以下三类 CR-LSP 能建立成功。

- Class-Type＝CT0，*setup-priority*＝6，*hold-priority*＝6。
- Class-Type＝CT0，*setup-priority*＝7，*hold-priority*＝6。
- Class-Type＝CT0，*setup-priority*＝7，*hold-priority*＝7。

说明 因为 CR-LSP 的建立优先级不能高于保持优先级，所以以上没有 *setup-priority*=6，*hold-priority*=7 的组合。

CT 和抢占优先级可以任意组合，因此理论上 TE-Class 共有 64 个，但华为设备最多支持手工配置 8 个 TE-Class。TE-Class 映射表是由一组 TE-Class 组成。建议 MPLS 网络中所有的 LSR 都配置相同的 TE-Class 映射表。如表 9-1 所示是设备中预先设置了缺省的 TE-Class 映射表，只包括了 CT0~CT3 这四种 CT 与抢占优先级间的映射关系（CT4~CT7 这四个 CT 没有配置缺省映射），并且分别与最低抢占优先级（7）和最高抢占优先级（0）进行了映射。

表 9-1　　缺省的 TE-Class 映射

TE-Class	CT	抢占优先级
TE-Class[0]	0	0
TE-Class[1]	1	0
TE-Class[2]	2	0
TE-Class[3]	3	0
TE-Class[4]	0	7
TE-Class[5]	1	7
TE-Class[6]	2	7
TE-Class[7]	3	7

9.1.4 DS-TE 中的带宽类型

在 DS-TE 隧道中，存在以下几种带宽类型，它们之间的关系如图 9-4 所示。

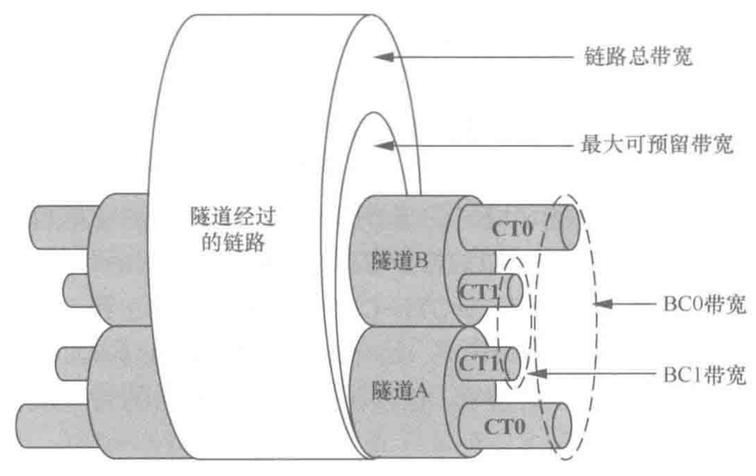


图 9-4　各种带宽之间的关系

■ 链路总带宽：物理链路所具有的总带宽。

■ 最大可预留带宽：本链路中可以预留给 MPLS TE 隧道使用的带宽，最大可预留带宽小于等于链路总带宽。

■ CT 带宽：**每条 DS-TE 隧道**中各类业务流量的预留带宽，用 CT*i* 表示。一条链路上可能建立多条 DS-TE 隧道，每条 DS-TE 隧道均有 CT*i*。

■ BC（Bandwidth Constraints，带宽约束）带宽：为该链路中**所有 DS-TE 隧道**的所有相同 CT*i* 预留的总带宽约束（用 BC*i* 表示），而 CT 带宽是具体 DS-TE 隧道中相同 CT*i* 的带宽约束。显然，BC 所包括的范围至少不会比 CT 所包括的范围小。

在隧道 A、隧道 B 中都包括了 CT0、CT1，它们都有自己的预留带宽 ，而 BC0 则包括隧道 A 和隧道 B 上的两个 CT0 预留的带宽总和，BC1 则包括隧道 A 和隧道 B 上的两个 CT1 预留的带宽总和。“最大可预留带宽”是在链路上设置的，链路上所有 DS-TE 隧道、所有 CT 所分配的预留带宽总和都不能超过总带宽。

9.1.5　DS-TE 带宽约束模型

带宽约束模型用来定义 BC 的最大数目，每个 BC 的带宽可被哪些 CT 使用，以及 CT 如何使用 BC 带宽。目前，IETF 定义了 MAM、RDM 和 Exended-MAM 三种带宽约束模型，下面具体介绍。

1. 最大分配模型（MAM，Maximum Allocation Model）

MAM 的 BC Mode ID 为 1，是将一个 BC 映射到链路上的一类 CT，CT 间不共享带宽，如图 9-5 所示。链路中所有相同 CT（CT*i*）的带宽总和不超过 BCi（$0 \leqslant i \leqslant 7$），所以 BC 的带宽总和不超过链路最大可预留带宽。

例如，假设某链路的带宽是 100Mbit/s，带宽模型为 MAM，且支持 3 个 CT（CT0、CT1 和 CT2）。BC0 为 20Mbit/s，用于承载 CT0（假设为 BE 流）；BC1 为 50Mbit/s，用于承载 CT1（假设为 AF 流）；BC2 为 30Mbit/s，用于承载 CT2（假设为 EF 流），则承载 BE 流的所有 CR-LSP 的带宽总和不能够超过 20Mbit/s；承载 AF 流的所有 CR-LSP 的带宽总和不能够超过 50Mbit/s；承载 EF 流的所有 CR-LSP 的带宽总和不能够超过 30Mbit/s。

MAM 的优点是不存在 CT 间的带宽抢占，缺点是可能存在带宽浪费，部分 CT 的带

宽利用率低。

2. 俄罗斯套娃模型（RDM，Russian Dolls Model）

RDM 允许 CT 间共享带宽，其 BC Mode ID 为 0。

RDM 的基本规则是：BC0≤链路的最大可预留带宽，BC*i*（0≤*i*≤7）≥各条 CR-LSP 中 CT*i*～CT7 的总带宽。也就是 *i* 编号的 BC 会大于等于自 *i* 编号开始，以后各编号 CT 的预留带宽总和。如 BC0≥CT0+CT1+CT2+CT3+CT4+CT5+CT6+CT7 预留带宽之和，同理 BC1 ≥ CT1+CT2+CT3+CT4+CT5+CT6+CT7 预留带宽之和，BC2 ≥ CT2+CT3+CT4+CT5+CT6+CT7 预留带宽之和……。此时各 BC 之间存在不同的包含关系，BC0 包含 BC1 的带宽，BC1 包含 BC2 的带宽，依次类推，最后 BC7 的带宽固定不变，如图 9-6 所示。

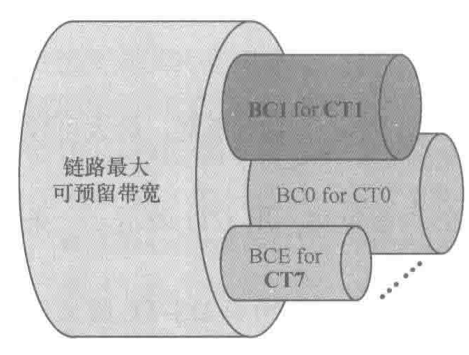


图 9-5 MAM 模型示意

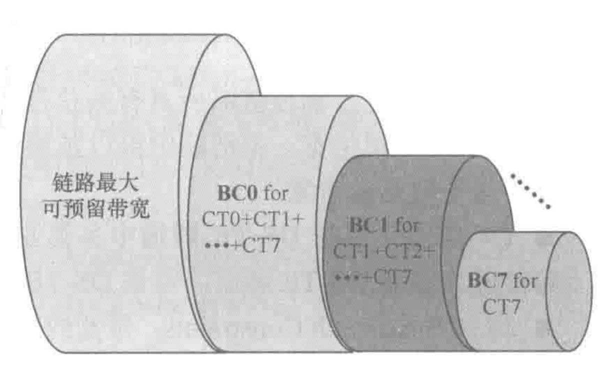


图 9-6 RDM 模型示意

这种模型与俄罗斯玩偶类似：一个大娃娃套一个小娃娃，小娃娃又套一个更小的娃娃，依次类推。从中可以看出，这与前面介绍的 BC*i* 仅是链路上各 CR-LSP 中 CT*i* 的预留带宽之和的定义不一样。

例如，假设某链路的带宽是 100Mbit/s，带宽模型为 RDM，且支持 3 个 CT（CT0、CT1 和 CT2）。其中，CT0 用于承载 BE 流，CT1 用于承载 AF 流，CT2 用于承载 EF 流。BC0 为 100Mbit/s，BC1 为 50Mbit/s，BC2 为 20Mbit/s。那么承载 EF 流的所有 LSP 带宽总和不超过 20Mbit/s，承载 EF 流和 AF 流的所有 LSP 带宽总和超过 50Mbit/s，所有 LSP 的带宽总和不超过 100Mbit/s。

RDM 模型允许存在 CT 间的带宽抢占。各 CT 的抢占关系是：如果 $0 \le m < n \le 7$，$0 \le i < j \le 7$，则优先级为 *m* 的 CT*i* 可以抢占优先级为 *n* 的 CT*i* 的带宽和优先级为 *n* 的 CT*j* 的带宽，比如，优先级为 3 的 CT0 可以抢占优先级为 5 的 CT0 带宽和优先级为 5 的 CT1 带宽。但是 CT*i* 的所有 LSP 带宽总和不超过 BC*i* 的带宽值。

RDM 模型的优点是可以有效利用带宽，缺点是优先级低的 CT 分类流量的带宽可能得不到保障，因为它们可以被更高优先级的 CT 业务流量长时间抢占。

3. 扩展的最大分配模型（Extended-MAM）

Extended-MAM 是为了支持下节将要介绍的 E-LSPu 差分服务方案而提出的一种带宽分配模式。Extended-MAM 的带宽分配基本规则与 MAM 模式类似，也是将一个 BC 映射到一个 CT，CT 间不共享带宽，其 BC Mode ID 为 254。但 Extended-MAM 比 MAM 模型多支持了 8 个隐式 CT，即 CT0 和 8 个抢占优先级（0～7）的组合，对应 TE-Class[8]～

TE-Class[15]，相当于对最低服务等级的 CT0 再根据不同的抢占优先级进行了细分，再分成了 8 个级别。

Extended-MAM 模型对 IGP 发布的 Unreserved Bandwidth Sub-TLV 及 Bandwidth Constraint Sub-TLV 的语义进行了新的阐释，Unreserved Bandwidth Sub-TLV 携带 CT0 对应其 8 个抢占优先级的 TE-Class 的未预留带宽，而用户指定的最多 8 个 TE-Class 的未预留带宽在通过 Bandwidth Constraint Sub-TLV 携带，使设备将支持的 TE-Class 扩展成了最多支持 16 个。

当配置了 Extended-MAM 模型的设备 A 作为 Transit 或 Egress 节点时，如果一非标准 DS-TE 设备 B 作为入节点创建有预留带宽的动态 CR-LSP，假使该 CR-LSP 的 TE-Class（<CT0，*priority*>，0≤*priority*≤7）未定义在设备 A 上用户指定的 TE-Class 映射表中，但是由于设备 A 配置的是 Extended-MAM 模型，所以该 CR-LSP 的创建请求也是合法的。

以上三种带宽约束模型的比较如表 9-2 所示。

表 9-2　带宽约束模型的比较

带宽约束模型间比较内容	RDM	MAM/Extended-MAM
BC 与 CT 映射关系	将 BC 映射到一个或多个 CT，不易管理	BC 与 CT 间一对一映射，易于理解和管理
带宽抢占	无法分隔不同 CT，需要抢占以保证为 CT 提供足够带宽	可以分隔不同 CT 并为 CT 提供有保证的带宽，不存在抢占
带宽利用率	有效带宽利用，因为不同 BC 间可带宽共享	可能造成带宽浪费，因为 BC 间不能带宽共享

9.1.6　DS-TE 差分服务方案

MPLS 与 Diff-Serv 都具有很好的可扩展性，而且处理过程也类似，如都是在网络边缘对报文进行分类（MPLS 基于标签，Diff-Serv 基于 DSCP 优先级值），在网络核心按标签，或 DSCP 映射的 PHB 进行转发。如果将 DSCP 优先级设置融入到 MPLS 标签的分配过程中，则 MPLS 标签将具备区分分组服务质量的能力。但是，在 IP 报头中的 DSCP 字段值对于 MPLS 的核心 LSR 设备（即 P 节点）是不可见的，因此必须存在某种机制让 DSCP 字段的值对 LSR 是可见的。根据将 IP 报文中的 Diff-Serv 信息通过标签传达给 LSR 的方式不同，RFC3270 定义了如下两种方案。

（1）L-LSP（Label-Only-Inferred-PSC LSP，仅由标签指示包交换能力的 LSP）

PSC 的全称为 Packet Switching Capable，即包交换能力。L-LSP 方案是在 LER 上将报文中的 DSCP 优先级值与一个 CR-LSP 建立映射（LER 对 IP 报头的 DSCP 字段是可见的），**使得同一 CR-LSP 上传输的报文将按相同的优先级进行处理，也将全部被分配到对应 DSCP 所映射的同一队列中进行调度**。这也就间接形成了 LSP 与队列之间的映射关系，所以后面的 LSR 可通过对应 LSP 的标签对报文进行队列调度，根据报文中的 EXP 优先级值进行丢弃选择（即 EXP 值决定 PHB）。

（2）E-LSP（EXP-Inferred-PSC LSP，由 EXP 指示包交换能力的 LSP）

这种方案是在 LER 上将报文中的 DSCP 优先级值映射到 EXP 优先级值，通过报文中的 EXP 优先级值为报文定义优先级别，一个 CR-LSP 最多可支持 8 个服务等级。这样

一来，**在一个 CR-LSP 中传输的报文可根据所携带的不同 EXP 值进入到不同的队列中进行调度**（EXP 与队列之间有映射关系）。同时根据 EXP 优先级值对报文进行丢弃选择（即 EXP 值决定 PHB），但 E-LSP 适用于支持不多于 8 个 PHB 的网络，因为 EXP 只有 8 个取值。

要想使不同业务流量在同一 CR-LSP 上传输时具有不同的优先级，就需要使用 E-LSP 方案，这也是 DS-TE 最主要的应用方案。目前华为 AR G3 系列路由器支持 E-LSP 方案，通过将 DSCP 或 EXP 映射到本地优先级（LP），实现 DSCP、EXP、LP 之间的相互映射。DSCP、EXP、LP 之间的缺省映射关系如表 9-3 所示。而本优先级缺省与队列优先级是一一对应的，这样一来，通过 DSCP 或 EXP 优先级与 LP 优先级的映射，间接地形成了 DSCP、EXP 优先级与队列优先级之间的映射关系。

表 9-3 DSCP、LP、EXP 缺省的映射关系

DSCP	LP	EXP
0～7	0	0
8～15	1	1
16～23	2	2
24～31	3	3
32～39	4	4
40～47	5	5
48～55	6	6
56～63	7	7

DS-TE 通过将不同的 LP 映射不同的 CT 上，使不同 CT 上传输的报文进行入不同的队列中，且可为每个 CT 单独分配资源，因此，DS-TE CR-LSP 是基于 CT 建立的。即 DS-TE 在路径计算过程中，需要将 CT 及每个 CT 可获得的带宽作为约束条件；在进行资源预留过程中，也需要考虑 CT 及其带宽需求。

9.1.7 DS-TE 模式及切换

华为设备目前支持 Non-IETF 模式和 IETF 两种 DS-TE 模式，主要区别就在于所支持的流量分类数不同，具体如下。

■ Non-IETF 模式：仅支持 2 种 CT（CT0 和 CT1），一种 CT 映射到一种流量类型，如将 AF 流量映射到 CT0，将 EF 流量映射到 CT1。当同一 VPN 中有多种流量类型时，仅支持其中 2 种流量分别定义为 CT0、CT1。Non-IETF 模式支持与 8 个优先级的组合，共 16 种 TE-Class，即最多可支持 16 种流量分类。

■ IETF 模式：IETF 定义的模式，支持 8 个 CT 和 8 个优先级的组合，共 64 种 TE-Class，即最多支持 64 种流量分类，但目前最多可配置 8 个 TE-Class。

Non-IETF 和 IETF 这两种 DS-TE 模式的详细比较如表 9-4 所示。

表 9-4 IETF 模式和 Non-IETF 模式的区别

比较项目	Non-IETF 模式	IETF 模式
带宽模型	支持 MAM 和 RDM	支持 RDM、MAM 和 Extended MAM
CT 类型	支持 CT0 和 CT1	支持 CT0～CT7

（续表）

比较项目	Non-IETF 模式	IETF 模式
BC 类型	支持 BC0 和 BC1	支持 BC0～BC7
TE-Class 映射表	可以配置 TE-Class 映射表，但不生效	支持配置和使用 TE-Class 映射表
IGP 消息	• 由 Unreserved Bandwidth sub-TLV 携带 CT0 分对应其 8 个优先级的 TE-Class 的未预留带宽，单位是 byte/s • 由 Unreserved Bandwidth for Class-Type 1（type 子字段值为 0x8001）sub-TLV 携带 CT1 对应其 8 个优先级的 TE-Class 的未预留带宽，单位是 byte/s	同时由 Unreserved Bandwidth sub-TLV 和 Bandwidth Constraints sub-TLV 携带 CT 信息。 • Unreserved Bandwidth sub-TLV： • 对于 RDM 和 MAM，携带 8 个 TE-Class 的未预留带宽（可以为用户指定），单位是 byte/s。 • 对于 Extended MAM，携带 CT0 分别与其对应的 8 个优先级的 TE-Class 的未预留带宽，单位是 byte/s。 • Bandwidth Constraints sub-TLV： • 对于 RDM 和 MAM，携带带宽模型信息及 BC 带宽。 • 对于 Extended MAM，携带带宽模型信息以及 8 个 TE-Class 的未预留带宽（可以为用户指定），单位是 byte/s
RSVP 消息	由 ADSPEC（通告说明）对象携带 CT 信息	• 单 CT：由 CLASSTYPE（服务分类）对象携带 CT 信息。 • 多 CT：由 EXTENDED_CLASSTYPE 对象携带 CT 信息

说明　为了实现 IETF 模式的 DS-TE，IETF 对 RSVP 进行了扩展：RFC4124 为 Path 消息定义了 CLASSTYPE 对象，用于携带 CT 类型；IETF 草案（draft-minei-diffserv-te-multi-class-02）为 Path 消息定义了 EXTENDED_CLASSTYPE 对象，用于携带 E-LSP 的 CT 类型信息。

华为 AR G3 系列路由器支持 Non-IETF 模式向 IETF 模式的切换，也支持 IETF 模式向 Non-IETF 模式的切换，具体处理方式如表 9-5 所示。

表 9-5　DS-TE 模式切换处理方式

比较项目	Non-IETF 模式→IETF 模式	IETF 模式→Non-IETF 模式
带宽模型变化情况	带宽模型不变	• Extended-MAM→MAM。 • RDM→RDM。 • MAM→MAM
带宽变化情况	BC0 和 BC1 的带宽值保持不变	除 BC0 和 BC1 保持不变外，其他 BC 的值被清零
TE-Class 映射表变化情况	如果已配置 TE-Class 映射表，则使用配置的 TE-Class 映射表，否则采用缺省的 TE-Class 映射表。缺省的 TE-Class 映射表请参见表 5-22	不使用 TE-Class 映射表。 • 如果用户配置了 TE-Class 映射表，则不删除 TE-Class 映射表。 • 如果用户没有配置 TE-Class 映射表，则删除缺省的 TE-Class 映射表

（续表）

比较项目	Non-IETF 模式→IETF 模式	IETF 模式→Non-IETF 模式
LSP 删除情况	在 Ingress 节点和 Transit 节点删除<CT,set-priority>组合和<CT,hold-priority>组合不在 TE-Class 映射表中的 LSP	在 Ingress 节点和 Transit 节点删除以下类型的 LSP： • 多 CT LSP。 • CT2～CT7 的单 CT LSP

9.1.8 DS-TE CR-LSP 建立和业务调度

在第 5 章介绍的 MPLS TE 中需要建立具有约束条件的 CR-LSP，此处介绍的 DS-TE 是 TE 的扩展，所采用的 LSP 也是有约束条件的，所以也需要建立 DS-TE CR-LSP，同时还可配置不同业务的相同或不同调度方式。

1. DS-TE CR-LSP 建立

DS-TE CR-LSP 的建立与 5.4.1 节介绍的 MPLS TE CR-LSP 的建立过程类似，也要用到 RSVP 的 Path 和 Resv 两种主要消息的传递，具体流程如图 9-7 所示。

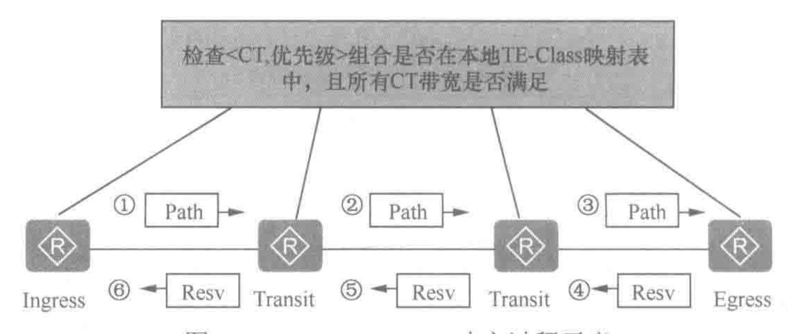


图 9-7 DS-TE CR-LSP 建立过程示意

DS-TE CR-LSP 建立与 TE CR-LSP 的建立过程中相比，主要区别体现在以下两个方面。

■ 在 DS-TE 中，RSVP 的 Path 消息除了包括第 5 章表 5-5 中所示的一系列对象外，还携带了所配置的 CT 信息。

■ 当沿途的 LSR 收到带有 CT 信息的 RSVP Path 消息后，除了要根据 Path 消息构建自己的 PSB，更新 Path 消息外，还会检查其中的<CT,优先级>组合是否在本地 TE-Class 映射表中，且所有 CT 带宽是否满足。如果这两个条件都满足，才会接受建立新的 LSP。

■ DS-TE CR-LSP 建立成功后，各节点会重新计算各优先级的每个 CT 可预留带宽。这些预留信息会反馈给 IGP，向网络中的其他节点通告。

2. DS-TE 业务调度

在业务调度方面，TE 隧道入节点的上行接口会根据所配置的或缺省的优先级映射，或 QoS 复杂流分类标记报文的本地优先级，LP（本地优先级）与 CT 服务类型一一对应，并携带到下行接口，下行接口再通过 HQoS（Hierarchical Quality of Service，分级 QoS）的多级调度功能实现 DS-TE 流量模型的带宽分配。

DS-TE 业务调度主要涉及以下几方面。

■ TE 隧道带宽预留

物理接口下预留所有 TE 隧道所需的带宽，通过单独的一级队列调度实现，可保证

TE带宽不被其他流量抢占。

■ TE隧道下CT业务的带宽保证

TE隧道支持多CT用以承载多种业务类型，其带宽在CR-LSP建立时进行端到端预留。设备在CR-LSP建立时，从TE预留带宽中分配各CT所需的带宽，遵从RDM/MAM等带宽模型约束。对于CT带宽通过CIR（承诺信息速率）保障。

■ CTs（不同LSP之间同一服务类型总和）之间的流量调度

不同CT对应不同的业务类型，对于高优先级业务（如语音），可采用PQ（Pripority Queue，优先级队列）调度。对于需要带宽保证的业务（如协议、数据等）可采用WFQ（Weighted Fair Queuing，加权公平队列）调度。因为CT与LP一一对应，LP又与队列一一对应，使得各CT与各队间有一相对应的映射关系，配置队列的调度模式，可以实现不同CT的差分服务。

华为AR G3系列路由器全局支持32个队列模板，可用于配置CT之间的调度关系。在缺省情况下，LP、CT和FQ（Fair Queue，公平队列）的映射关系如表9-6所示，序号方面都是一一对应的关系，CT之间的调度方式均为WFQ。

表9-6　LP、CT和FQ（公平队列）的缺省映射关系

本地优先级LP	CT类型	FQ队列	调度方式（可配置）
7 (CS7)	CT7	7	WFQ
6 (CS6)	CT6	6	WFQ
5 (EF)	CT5	5	WFQ
4 (AF4)	CT4	4	WFQ
3 (AF3)	CT3	3	WFQ
2 (AF2)	CT2	2	WFQ
1 (AF1)	CT1	1	WFQ
0 (BE)	CT0	0	WFQ

9.2 静态DS-TE隧道配置与管理

DS-TE隧道就是在TE隧道基础上配置了QoS功能，实现了不同业务分类，并可为不同业务提供服务优先级和预留带宽保证。所以，DS-TE隧道建立的前提也是先把MPLS TE隧道建立好，然后再配置所需的QoS功能，就建立好DS-TE隧道了。

DS-TE隧道与MPLS TE隧道一样，也可以静态建立或动态方式建立。本节先介绍静态DS-TE隧道的配置方法。

9.2.1 静态DS-TE的配置任务

配置静态DS-TE隧道的过程比较简单，手工分配标签，不使用信令协议，不需要交互控制报文，因此消耗资源比较小，但静态DS-TE隧道只支持单CT，即一条静态CR-LSP上只能承载一类业务。

静态DS-TE所包括的配置任务比较多，除了最基本的MPLS TE能力的使能、MPLS TE Tunnel接口配置外，更多的是DS-TE隧道所特有的、为不同类型业务流量提供的QoS

服务配置，包括 DS-TE 工作模式、带宽约束模型、链路带宽和各 CT 可预留带宽指定、优先级映射、调度方式等，具体如下。

（1）使能 MPLS TE

这项配置任务的配置方法与第 5 章的 5.6.1 节介绍的配置方法完全一样，参见即可。

（2）配置 MPLS TE 隧道接口

这项配置任务的配置方法与第 5 章的 5.6.2 节介绍的配置方法完全一样，参见即可。

（3）配置 DS-TE 模式

这是全局配置本设备工作在 Non-IETF 或 IETF 模式，将作用于设备上所创建的所有 DS-TE 隧道。需根据所需区分的业务类型多少来选择，多于 2 种时必须选择 IETF 模式。

（4）配置 DS-TE 带宽约束模型

这是全局配置设备中 DS-TE 隧道选择 Extend-MAM、MAM 或 RDM 的带宽约束模型，作用于本设备上所创建的所有 DS-TE 隧道。需根据各 CT 间是否可抢占带宽资源来选择。

（5）（可选）配置 TE-Class 映射表

这是一项全局配置任务，当需要修改缺省的 CT 与抢占优先级映射关系时才需要配置，但一旦修改则同时适用于所有本设备上创建的 DS-TE 隧道。

（6）配置链路的带宽

这是一项针对具体链路的配置任务，配置本地链路上最大可预留带宽，同时指定在本地链路上各 DS-TE 隧道中各业务类型（对应不同 CT）可分配的带宽。但一条静态 CR-LSP 只能配置一个 CT，只能承载一类业务流量。

（7）配置静态 CR-LSP 并指定带宽

这是在入节点基于具体 Tunnel 接口创建静态 CR-LSP，在中间节点和出节点基于隧道名称创建静态 CR-LSP，并为所需 CT 指定预留的带宽（出节点上不用预留带宽）。

（8）配置接口信任的报文优先级

这是在具体的 DS-TE 隧道所使用的物理接口来配置的，在入节点上配置信任 DSCP 优先级，以对报文进行分类；在中间节点配置信任 EXP 优先级，使报文按 EXP 优先级转发；在出节点信任 DSCP 或 EXP 优先级，使报文按照 DSCP 优先级或 EXP 优先级转发。

（9）（可选）配置 CT 与业务类型的映射关系以及调度方式

配置 CT 与业务类型映射也是一项全局配置，作用于本设备上所有建立的 DS-TE 隧道，其目的是把不同业务类型流量分配到不同的队列中进行转发，主要是通过调整 DSCP、EXP 优先级与 LP 优先级，或者 DSCP 与 DSCP、EXP 与 EXP 优先级之间的映射关系来实现的。队列调度方式是在 DS-TE 隧道各节点的出接口进行配置的。

下面仅就后面 7 项配置任务的配置方法进行介绍。

在配置静态 DS-TE 隧道之前，仍需要完成基本的 MPLS TE 隧道建立所需的配置任务，具体包括。

- 配置 IGP 路由协议，使各节点间的 IP 路由可达。
- 配置各 LSR 节点的 LSR-ID，以及全局和各 LSR 节点的接口 MPLS 能力。

9.2.2 配置 DS-TE 模式和带宽约束模型

1. 配置 DS-TE 模式

通过配置 DS-TE 模式，可以将 DS-TE 隧道设置为 Non-IETF 或 IETF 模式，使设备可以支持不同数量的业务类型和带宽约束模型。

■ Non-IETF（非标准）模式：支持两种 CT（CT0 和 CT1）和 8 种抢占优先级（0～7），带宽约束模型支持 RDM 和 MAM。

■ IETF（标准）模式：支持 8 种 CT（CT0～CT7）和 8 种抢占优先级（0～7）。带宽约束支持 RDM、MAM 和 Extended MAM 三种模型。

需在 DS-TE 隧道各节点进行表 9-7 所示的配置。

表 9-7 DS-TE 模式的配置步骤

步骤	命令	说明
1	**system-view** 例如：<Huawei> **system-view**	进入系统视图
2	**mpls** 例如：[Huawei] **mpls**	进入 MPLS 视图
3	**mpls te ds-te mode** { **ietf** | **non-ietf** } 例如：[Huawei-mpls] **mpls te ds-te mode ietf**	配置 DS-TE 模式。命令中的选项说明如下。 • **Ietf**：二选一选项，指定采用标准 DS-TE 模式。 • **non-ietf**：二选一选项，指定采用非标准 DS-TE 模式。 缺省情况下，DS-TE 模式为非标准（Non-IETF）模式，可用 **undo mpls te ds-te mode** 命令恢复缺省配置

华为设备支持非标准（Non-IETF）模式向标准（IETF）模式的切换，也支持 IETF 向 Non-IETF 模式的切换。

2. 配置 DS-TE 带宽约束模型

如果网络允许 CT 之间的带宽资源抢占，则建议选择 RDM 模型，以实现带宽的有效利用，否则建议选择 MAM 或 Extended-MAM 模型。

带宽约束模型需在 DS-TE 隧道各节点进行配置，具体步骤如表 9-8 所示。

表 9-8 DS-TE 带宽约束模型的配置步骤

步骤	命令	说明
1	**system-view** 例如：<Huawei> **system-view**	进入系统视图
2	**mpls** 例如：[Huawei] **mpls**	进入 MPLS 视图
3	**mpls te ds-te bcm** { **extend-mam** | **mam** | **rdm** } 例如：[Huawei-mpls] **mpls te ds-te bcm mam**	配置 DS-TE 带宽约束模型。命令中的选项说明如下。 • **extend-mam**：多选一选项，设置 DS-TE 的带宽约束模型为 Extended-MAM 模型。当 DS-TE 模式为 Non-IETF 模式时，不可选择。 • **mam**：多选一选项，设置 DS-TE 的带宽约束模型为最大分配模型 MAM。 • **rdm**：多选一选项，设置 DS-TE 的带宽约束模型为俄罗斯套娃模型 RDM

（续表）

步骤	命令	说明
3	**mpls te ds-te bcm** { **extend-mam** | **mam** | **rdm** } 例如：[Huawei-mpls] **mpls te ds-te bcm mam**	当非标准 DS-TE 模式切换为标准 DS-TE 模式时，带宽模型不变。标准 DS-TE 模式切换为非标准 DS-TE 模式时，切换前如果是 Extended-MAM 模型，则切换后，系统自动将其改为 MAM 模型。其他模型切换前后不发生变化。 缺省情况下，DS-TE 带宽约束模型为 RDM，可用 **undo mpls te ds-te bcm** 命令恢复缺省配置

9.2.3 配置 TE-Class 映射表和链路带宽

1. 配置 TE- Class 映射表

对于非标准 DS-TE 模式，无需此步骤配置，即使配置了也不会生效。对于标准 DS-TE 模式，需要预先规划好 TE-Class 映射表。建议整个 DS-TE 域配置相同的 TE-Class 映射表，否则 LSP 可能不能正确建立。

配置 TE-Class 映射表时需了解以下内容。

- TE-Class 映射表的配置在每台设备上是唯一的。
- TE-Class 映射表是全局概念，TE-Class 映射表应用到该 LSR 的所有 DS-TE 隧道中。

说明 TE-Class 是 CT 和优先级 *priority* 的组合<CT，*priority*>。这里的 *priority* 是指 CR-LSP 的抢占优先级，而不是 MPLS 首部的 EXP 值。抢占优先级的取值范围是 0～7，**数值越小，优先级越高**。只有当一条 CR-LSP 的 CT 和建立优先级的组合<CT，*setup-priority*>以及 CT 和保持优先级的组合<CT，*hold-priority*>同时存在于 TE-Class 映射表中，该 CR-LSP 才能建立成功。CR-LSP 的建立优先级不能高于保持优先级。

- MAM 和 Extended MAM 模型中，高优先级的 CT 只能抢占相同类型值的 CT 的带宽；不同 CT 之间不发生带宽抢占。
- RDM 模型中，各 CT 之间的带宽抢占同时受抢占优先级及对应 BC 的限制。假设 *m*、*n* 为抢占优先级值，*i*、*j* 为 CT 类型值，其中 $0 \leq m < n \leq 7$，且 $0 \leq i < j \leq 7$，则：
- 优先级为 *m* 的 CT*i* 可以抢占优先级为 *n* 的 CT*i* 带宽，或优先级为 *n* 的 CT*j* 带宽。
- 所有 CT*i* 的带宽总和≤BC*i* 的带宽值。
- 只有当一条 CR-LSP 的所有 CT 的带宽都满足要求，抢占才能发生，待建立的 CR-LSP 才能建立成功。

标准 DS-TE 模式的 TE-Class 映射需在 DS-TE 隧道各节点进行表 9-9 所示的配置。

表 9-9 TE-Class 映射表的配置步骤

步骤	命令	说明
1	**system-view** 例如：<Huawei> **system-view**	进入系统视图
2	**te-class-mapping** 例如：[Huawei] **te-class-mapping**	创建 TE-Class 映射表并进入 TE-Class Mapping 视图

（续表）

步骤	命令	说明
	选择如下的一条或多条命令，配置各个 TE-Class	
3	**te-class0 class-type** { **ct0** | **ct1** | **ct2** | **ct3** | **ct4** | **ct5** | **ct6** | **ct7** } **priority** *priority* [**description** *description-info*] 例如：[Huawei-te-class-mapping] **te-class0 class-type ct0 priority** 0 **description** For-EF	配置 TE-Class0。命令中参数和选项说明如下。 ● **ct0**～**ct7**：多选一选项，指定业务类型值。 ● *priority*：指定抢占优先级，整数形式，取值范围是 0～7。数值越小，优先级越高。 ● **description** *description-info*：可选参数，配置指定该 TE-Class 的描述信息。 缺省情况下，没有配置 TE-Class，可用 **undo te-class0** 命令删除一个 TE-Class0 的配置
	te-class1 class-type { **ct0** | **ct1** | **ct2** | **ct3** | **ct4** | **ct5** | **ct6** | **ct7** } **priority** *priority* [**description** *description-info*]	配置 TE-Class1。其他说明参见 **te-class0 class-type** 命令介绍
	te-class2 class-type { **ct0** | **ct1** | **ct2** | **ct3** | **ct4** | **ct5** | **ct6** | **ct7** } **priority** *priority* [**description** *description-info*]	配置 TE-Class2。其他说明参见 **te-class0 class-type** 命令介绍
	te-class3 class-type { **ct0** | **ct1** | **ct2** | **ct3** | **ct4** | **ct5** | **ct6** | **ct7** } **priority** *priority* [**description** *description-info*]	配置 TE-Class3。其他说明参见 **te-class0 class-type** 命令介绍
	te-class4 class-type { **ct0** | **ct1** | **ct2** | **ct3** | **ct4** | **ct5** | **ct6** | **ct7** } **priority** *priority* [**description** *description-info*]	配置 TE-Class4。其他说明参见 **te-class0 class-type** 命令介绍
	te-class5 class-type { **ct0** | **ct1** | **ct2** | **ct3** | **ct4** | **ct5** | **ct6** | **ct7** } **priority priority** [**description description-info**]	配置 TE-Class5。其他说明参见 **te-class0 class-type** 命令介绍
	te-class6 class-type { **ct0** | **ct1** | **ct2** | **ct3** | **ct4** | **ct5** | **ct6** | **ct7** } **priority priority** [**description description-info**]	配置 TE-Class6。其他说明参见 **te-class0 class-type** 命令介绍
	te-class7 class-type { **ct0** | **ct1** | **ct2** | **ct3** | **ct4** | **ct5** | **ct6** | **ct7** } **priority priority** [**description description-info**]	配置 TE-Class7。其他说明参见 **te-class0 class-type** 命令介绍

2. 配置链路的带宽

通过配置链路的带宽可以限定 DS-TE 隧道的带宽。带宽约束模型不同，链路的可预留带宽与各 BC 的带宽之间的关系不同（*max-reservable-bandwidth* 是指链路最大可预留带宽），具体如下。

■ RDM 模型：*max-reservable-bandwidth*≥*bc0-bw-value*≥*bc1-bw-value*≥*bc2-bw-value*≥*bc3-bw-value*≥*bc4-bw-value*≥*bc5-bw-value*≥*bc6-bw-value*≥*bc7-bw-value*。

■ MAM 模型：*max-reservable-bandwidth*≥*bc0-bw-value*＋*bc1-bw-value*＋*bc2-bw-value*＋*bc3-bw-value*＋*bc4-bw-value*＋*bc5-bw-value*＋*bc6-bw-value*＋*bc7-bw-value*。

■ Extended-MAM 模型：同 MAM 模型。

如需要进行精确的带宽控制，则需要配置链路上的 BC*i* 带宽值不小于经过该链路的所有 DS-TE 隧道上使用该 BC 带宽的 CT*i*（0≤*i*≤7）的带宽总和的 125%。比如。

■ MAM/Extended-MAM 模型中，BC*i* 带宽值≥CT*i* 带宽值× 125%（0≤*i*≤7）

■ RDM 模型中，BCi 带宽值≥CT*i*～CT7 的总带宽值× 125%（0≤*i*≤7）

需在 DS-TE 隧道各节点的**出接口**上进行表 9-10 所示配置。

表 9-10　　链路带宽的配置步骤

步骤	命令	说明
1	**system-view** 例如：<Huawei> **system-view**	进入系统视图
2	**interface** *interface-type interface-number* 例如：[Huawei] **interface** gigabitethernet 1/0/0	进入链路的出接口的视图
3	**mpls te bandwidth max-reservable-bandwidth** *bw-value* 例如：[Huawei-GigabitEthernet1/0/0] **mpls te bandwidth max-reservable-bandwidth** 10000	配置链路为 DS-TE 隧道可预留的最大带宽，整数形式，取值范围是 0～4000000000，单位是 kbit/s。缺省值是 0 缺省情况下，没有配置链路的最大可预留带宽，可用 **undo mpls te bandwidth max-reservable-bandwidth** 命令恢复系统缺省配置
4	**mpls te bandwidth** { **bc0** *bc0-bw-value* | **bc1** *bc1-bw-value* }* 例如：[Huawei-GigabitEthernet1/0/0] **mpls te bandwidth bc0** 1000	（二选一）非标准（Non-IETF）模式下配置链路的带宽。可多选参数 **bc0** *bc0-bw-value* | **bc1** *bc1-bw-value* 分别用来设置 BC0、BC1 的带宽，整数形式，取值范围是 1～4000000000，单位是 kbit/s。缺省值是 1kbit/s。 【说明】如果要修改 BC 带宽值，可重新配置该命令，最后一次的配置会覆盖之前的配置。不允许将 BC 带宽值修改成小于已经分配给 BC 的带宽值。例如，BC0 已有 10Mbit/s 的带宽被分配，则 BC0 的带宽只能修改成大于等于 10Mbit/s。 缺省情况下，没有配置链路的 BC 带宽，可用 **undo mpls te bandwidth** 命令恢复系统缺省配置
	mpls te bandwidth { **bc0** *bc0-bw-value* | **bc1** *bc1-bw-value* | **bc2** *bc2-bw-value* | **bc3** *bc3-bw-value* | **bc4** *bc4-bw-value* | **bc5** *bc5-bw-value* | **bc6** *bc6-bw-value* | **bc7** *bc7-bw-value* }* 例如：[Huawei-GigabitEthernet1/0/0] **mpls te bandwidth bc5** 1000	（二选一）标准（IETF）模式下配置链路的带宽。命令中的参数与 Non-IETF 模式下的 **mpls te bandwidth** 命令中的参数取值范围一样，只不过，在 IETF 模式下，可以分别为 BC0~BC7 配置链路带宽。其他说明参见 IETF 模式下 **mpls te bandwidth** 命令的说明

9.2.4 配置静态 CR-LSP 并指定带宽

配置静态 DS-TE 隧道时，需要在静态 DS-TE 隧道入节点、中间节点和出节点手工配置静态 CR-LSP。当没有中间节点时，可以不必配置中间节点的静态 CR-LSP。

1. 在 Ingress 节点上的配置

在系统视图下执行 **static-cr-lsp ingress** { **tunnel-interface tunnel** *interface-number* | *tunnel-name* } **destination** *destination-address* { **nexthop** *next-hop-address* | **outgoing-interface** *interface-type interface-number* }* **out-label** *out-label* **bandwidth** [**ct0** | **ct1** | **ct2** | **ct3** | **ct4** | **ct5** | **ct6** | **ct7**] *bandwidth* 命令，配置入节点的静态 CR-LSP，并指定 CT 及其带宽值。命令中的参数和选项说明如下。

■ **tunnel** *interface-number*：二选一参数，指定静态 CR-LSP 的隧道接口的编号，格

式为“槽位号/卡号/端口号”，槽位号、卡号均为整数形式，取值与设备有关，端口号为整数形式。

■ *tunnel-name*：二选一参数，指定静态 CR-LSP 隧道的名称，字符串形式，**区分大小写**，不支持空格和缩写，长度范围是 1～19，必须与命令 **interface tunnel** *interface-number* 创建的隧道接口名称一致。假设使用 **interface Tunnel** 0/0/1 命令为静态 CR-LSP 创建了一个 Tunnel 接口，则入节点中的该参数应该写作“Tunnel0/0/1”，否则隧道将不能正确建立。中间节点和出节点无此限制。

■ **destination** *destination-address*：指定静态 CR-LSP 的目的地址，通常为 Egress 节点的 Loopback 接口 IP 地址。

■ **nexthop** *next-hop-address*：可多选参数，指定静态 CR-LSP 下一跳地址。下一跳或出接口由入节点到出节点的路由决定。如果 LSP 出接口为以太网类型，必须配置 **nexthop** *next-hop-address* 参数以保证 LSP 的正常转发。

注意 如果在配置静态 CR-LSP 时指定了下一跳，则在配置 IP 静态路由时也必须指定下一跳，否则不能建立静态 CR-LSP。

■ **outgoing-interface** *interface-type interface-number*：可多选参数，指定出接口类型和编号，只有点到点链路才能选择配置出接口。

■ **out-label** *out-label*：指定出标签的值，整数形式，取值范围是 16～1048575。上游节点的出标签也是下游节点的入标签，这点与 LDP LSP 是一样的。

■ **bandwidth** [**ct0** | **ct1** | **ct2** | **ct3** | **ct4** | **ct5** | **ct6** | **ct7**] *bandwidth*：用来指定静态 CR-LSP 的 CT0～CT7 带宽值，整数形式，取值范围是 0～4000000000，单位为 kbit/s，缺省值为 0。

注意 静态 CR-LSP 只支持单 CT，即对于标准 DS-TE，配置该命令时仅可以选择 CT0～CT7 中的任意一个取值；对于非标准 DS-TE，配置该命令时仅可以选择 CT0 和 CT1 中的任意一个取值。

静态 CR-LSP 具有最高的抢占优先级（优先级数值为 0），所以不能被其他 CR-LSP 抢占，但静态 CR-LSP 创建时也不抢占其他 LSP 的资源，尽管他具有最高的抢占优先级。

隧道的带宽不能超过链路最大可预留带宽。此外，同一个节点上无论使用哪种带宽约束模型，所有 CT*i* 的带宽总和不超过 BC*i* 的带宽值（$0 \leqslant i \leqslant 7$），即 CT*i* 只能使用 BC*i* 的带宽。例如，某 PE 节点的 BC1 带宽值为 x，该节点一共有两条 CT1 的静态 CR-LSP，带宽分别为 y 和 z，则 $(y+z) \leqslant x$。

缺省情况下，没有在入口节点配置静态 CR-LSP，可用 **undo static-cr-lsp ingress** { **tunnel-interface tunnel** *interface-number* | *tunnel-name* }命令在入口节点删除配置的静态 CR-LSP。但如果要对除 Tunnel 接口外的其他参数进行修改，可直接重新执行本命令进行配置，不用删除原来的 CR-LSP。

例如要在入口节点配置静态 CR-LSP。名字为 Tunnel0/0/1，目的 IP 地址为 10.1.3.1，下一跳 IP 地址为 10.1.1.2，出标签为 237，从 CT1 获取带宽，所需要的带宽 20kbit/s。

```
<Huawei> system-view
[Huawei] static-cr-lsp ingress Tunnel0/0/1 destination 10.1.3.1 nexthop 10.1.1.2 out-label 237 bandwidth ct1 20
```

2. 在 Transit 节点上的配置

在系统视图下执行 **static-cr-lsp transit** *lsp-name* [**incoming-interface** *interface-type interface-number*] **in-label** *in-label* { **nexthop** *next-hop-address* | **outgoing-interface** *interface-type interface-number* } * **out-label** *out-label* **bandwidth** [**ct0** | **ct1** | **ct2** | **ct3** | **ct4** | **ct5** | **ct6** | **ct7**] *bandwidth* 命令，配置中间节点的静态 CR-LSP，并指定 CT 及其带宽值。命令中的参数和选项说明如下。

■ *lsp-name*：指定 CR-LSP 隧道的名字，符串形式，区分大小写，不支持空格，长度范围是 1～19。当输入的字符串两端使用双引号时，可在字符串中输入空格，名称取值没有限制，但不能与该节点上已存在的名称相同。为了清晰，可以使用此静态 CR-LSP 的 MPLS TE 隧道的接口名称，如 Tunnel0/0/1。

■ **incoming-interface** *interface-type interface-number*：可选参数，指定 CR-LSP 的入接口。

■ **in-label** *in-label*：指定入标签的值，整数形式，取值范围是 16～1023。

■ **nexthop** *next-hop-address*：可多选参数，指定下一跳 IP 地址。如果 LSP 出接口为以太网类型，必须配置 **nexthop** *next-hop-address* 参数以保证 LSP 的正常转发。

■ **outgoing-interface** *interface-type interface-number*：可多选参数，指定出接口名称。

■ **out-label** *out-label*：指定出标签的值，整数形式，取值范围是 16～1048575。

注意 因为是出标签与入标签的取值范围不一样，为了确保上游节点的出标签与下游节点的入标签保持一致，需要在配置出标签时，必须不能超过入标签的取值范围 16～1023。

■ **bandwidth** [**ct0** | **ct1** | **ct2** | **ct3** | **ct4** | **ct5** | **ct6** | **ct7**] *bandwidth*：用于指定静态 CR-LSP 的 CT0～CT7 带宽值的参数，带宽取值为整数形式，取值范围是 0～4000000000，单位为 kbit/s，缺省值为 0。

缺省情况下，没有在转发节点配置静态 CR-LSP，可用 **undo static-cr-lsp transit** *lsp-name* 命令在转发节点删除指定的静态 CR-LSP。但如果要对除 CR-LSP 隧道名称外的其他参数进行修改，可直接重新执行本命令进行配置，不用删除原来的 CR-LSP。

例如在中间节点配置静态 CR-LSP，名字为 tunnel39，入接口为 GE1/0/0，入口标签为 123，出接口为 GE2/0/0，出口标签为 253，CR-LSP 所需 CT0 带宽为 20kbit/s。

```
<Huawei> system-view
[Huawei] static-cr-lsp transit tunnel39 incoming-interface gigabitethernet 1/0/0 in-label 123 outgoing-interface gigabitethernet 2/0/0 out-label 253 bandwidth ct0 20
```

3. 在 Egress 节点上的配置

在系统视图下执行 **static-cr-lsp egress** *lsp-name* [**incoming-interface** *interface-type interface-number*] **in-label** *in-label* [**lsrid** *ingress-lsr-id* **tunnel-id** *tunnel-id*]，配置出节点的静态 CR-LSP。命令中的参数说明如下。

■ *lsp-name*：指定 CR-LSP 隧道的名称，取值名称也没有限制，但不能与该节点上已存在的名称相同。为了清晰，可以使用此静态 CR-LSP 的 MPLS TE 隧道的接口名称，如 Tunnel0/0/1。

- **incoming-interface** *interface-type interface-number*：可选参数，指定入接口。
- **in-label** *in-label*：指定入标签的值，整数形式，取值范围是 16～1023。
- **lsrid** *ingress-lsr-id*：可选参数，指定入节点的 LSR ID。
- **tunnel-id** *tunnel-id*：可选参数，指定隧道标识，整数形式，取值范围是 1～65535。

缺省情况下，没有在出口节点配置静态 CR-LSP，可用 **undo static-cr-lsp egress** *lsp-name* 命令在出口节点删除配置的静态 CR-LSP。同样，如果要对除 CR-LSP 隧道名称外的其他参数进行修改，可直接重新执行本命令进行配置，不用删除原来的 CR-LSP。

如在出节点配置静态 CR-LSP，名字为 tunnel34，入接口是 GE1/0/0，入口标签是 233。

```
<Huawei> system-view
[Huawei] static-cr-lsp egress tunnel34 incoming-interface gigabitethernet 1/0/0 in-label 233
```

9.2.5　配置接口信任的报文优先级

创建 DS-TE 隧道后，需要在入接口配置信任的报文优先级，从而对不同业务流量提供有差别的 QoS 服务质量。

信任的报文优先级配置都是在入接口上进行配置的，在入节点、中间节点和出节点上要配置的信任的优先级类型不一样。

- 在入节点的入接口上要通过 **trust dscp** 命令信任报文中的 DSCP 优先级，配置对报文按照 DSCP 优先级进行映射。
- 在中间节点的入接口上要通过 **trust exp** 命令信任报文中的 MPLS EXP 优先级，配置对报文按照 MPLS EXP 优先级进行映射。
- 在出节点的入接口上，如果配置了倒数第二跳弹出（PHP）功能后，入接口接收的报文为 IP 报文，此时要通过 **trust dscp** 命令信任报文中的 DSCP 优先级，配置对报文按照 DSCP 优先级进行映射；如果没有配置 PHP 功能，入接口接收的报文为 MPLS 报文，此时要通过 **trust exp** 命令信任报文中的 EXP 优先级，配置对报文按照 MPLS EXP 优先级进行映射。

9.2.6　配置 CT 与业务类型的映射关系以及调度方式

建立 DS-TE 隧道后，还需要配置 CT 与业务类型的映射关系，同时还可以配置各 CT 对应的调度方式。CT 与 LP 是一一对应的关系，通过调整 DSCP-LP 映射关系和 EXP-LP 的映射关系，可以实现 DSCP、LP、EXP 之间的映射关系调整，可实现整个 DS-TE 隧道的各 CT 与具有不同优先级的业务类型之间的映射。DSCP、LP、EXP 之间的缺省映射关系参见 9.1.6 节表 9-3。CT、LP、FQ 队列和调度方式的缺省映射关系参见 9.1.8 节的表 9-6。

华为设备目前还支持 DSCP-DSCP 映射关系和 EXP-EXP 映射关系调整，可根据实际规划进行灵活配置。

1. 调整 CT 与业务类型映射关系

可根据实际的优先级映射需要，需在 DS-TE 隧道各节点进行表 9-11 所示的配置，但并不一定需要全面修改这些优先级之间的映射关系，仅当某两个优先级之间的缺省映射关系需要修改时才需要选择配置。建议进行整体规划，配置相同的映射关系。

表 9-11 调整 CT 与业务类型映射关系的配置步骤

步骤	命令	说明
1	**system-view** 例如：<Huawei> **system-view**	进入系统视图
调整 DSCP-LP 映射关系		
2	**qos map-table dscp-lp** 例如：[Huawei] **qos map-table dscp-lp**	进入 dscp-lp 视图，即从 DSCP 到本地优先级的映射视图
3	**input** { *input-value1* [**to** *input-value2*] } &<1-10> **output** *output-value* 例如：[Huawei-maptbl-dscp-lp] **input** 0 **to** 10 **output** 0	配置 DSCP 和 LP 之间的映射关系。命令中的参数说明如下。 • *input-value1*：指定输入的起始 DSCP 值，整数形式，取值范围是 0～63。 • *input-value2*：可选参数，指定输入的终止 DSCP 值，整数形式，取值范围是 0～63。 • &<1-10>：表示前面的 *input-value1* [to *input-value2*] 参数最多可以有 10 组。 • *output-value*：指定输出的本地优先级值，整数形式，取值范围是 0～7。 缺省的 DSCP 到 LP 的映射关系如表 5-32 所示，可用 **undo input** { *input-value1* [to *input-value2*] } &<1-10> **output** *output-value* 命令删除指定的映射
调整 EXP-LP 映射关系		
2	**qos map-table exp-lp** 例如：[Huawei] **qos map-table exp-lp**	进入 exp-lp 视图，即从 EXP 到本地优先级的映射视图
3	**input** { *input-value1* [**to** *input-value2*] } &<1-10> **output** *output-value* 例如：[Huawei-maptbl-exp-lp] **input** 0 **output** 7	配置 EXP 和 LP 之间的映射关系。命令中的参数说明如下。 • *input-value1*：指定输入的起始 EXP 值，整数形式，取值范围是 0～7。 • *input-value2*：可选参数，指定输入的终止 EXP 值，整数形式，取值范围是 0～7。 • &<1-10>：表示前面的 *input-value1* [to *input-value2*] 参数最多可以有 10 组。 • *output-value*：指定输出的本地优先级值，整数形式，取值范围是 0～7。 缺省的 EXP 到 LP 的映射关系如表 5-32 所示，可用 **undo input** { *input-value1* [**to** *input-value2*] } &<1-10> **output** *output-value* 命令删除指定的映射
调整 DSCP-DSCP 映射关系		
2	**qos map-table dscp-dscp** 例如：[Huawei] **qos map-table dscp-dscp**	进入 dscp-dscp 视图，即从 DSCP 到 DSCP 的映射视图
3	**input** { *input-value1* [**to** *input-value2*] } &<1-10> **output** *output-value* 例如：[Huawei-maptbl-dscp-dscp] **input** 0 **to** 10 **output** 0	配置 DSCP 和 DSCP 之间的映射关系。命令中的 *input-value1*、*input-value2* 和 *output-value* 参数均为 DSCP 优先绩级值，整数形式，取值范围均是 0～63

（续表）

步骤	命令	说明
调整 EXP-EXP 映射关系		
2	**qos map-table exp-exp** 例如：[Huawei] **qos map-table exp-exp**	进入 exp-exp 视图，即从 MPLS EXP 到 MPLS EXP 优先级的映射视图
3	**input** { *input-value1* [**to** *input-value2*] } <1-10> **output** *output-value* 例如：[Huawei-maptbl-exp-exp] **input** 0 **to** 3 **output** 0	配置 EXP 和 EXP 之间的映射关系。命令中的 *input-value1*、*input-value2* 和 *output-value* 参数均为 EXP 优先绩级值，整数形式，取值范围均是 0～7

2. 配置 CT 调度方式

因为 CT 与 LP 一一对应，LP 又与队列一一对应，故通过配置队列的调度模式可以实现不同 CT 的差分服务。各 CT 的调度方式也有缺省值，参见 9.1.8 节的表 9-6，如需修改，则需在 DS-TE 隧道各节点的出接口上进行表 9-12 所示配置。

表 9-12　　CT 调度方式的配置步骤

步骤	命令	说明
1	**system-view** 例如：<Huawei> **system-view**	进入系统视图
2	**qos queue-profile** *queue-profile-name* 例如：[Huawei] **qos queue-profile** profile1	创建队列模板并进入队列模板视。参数 *queue-profile-name* 用来指定队列模板名称，字符串形式，不支持空格，区分大小写，长度范围是 1～31。 缺省情况下，系统中没有队列模板，可用 **undo qos queue-profile** *queue-profile-name* 命令删除已创建的指定队列模板。如果要删除的队列模板已经被接口绑定，需要先在相应的接口视图下执行 **undo qos queue-profile** 命令删除该模板在接口下的应用，再在系统视图下执行 **undo qos queue-profile** *queue-profile-name* 删除队列模板
3	**schedule** { **pq** *start-queue-index* [**to** *end-queue-index*] | **wfq** *start-queue-index* [**to** *end-queue-index*] }* 例如：[Huawei-qos-queue-profile-profile1] **schedule pq** 0 **to** 3	在队列模板中为指定队列配置调度模式。命令中的参数说明如下。 • **pq**：可多选选项，指定采用严格优先级调度模式。 • **wfq**：可多选选项，指定采用加权公平调度模式。 • *start-queue-index* [**to** *end-queue-index*：指定配置队列调度关系的队列的索引。其中：*start-queue-index* 表示配置队列调度关系的第一个队列，*end-queue-index* 表示配置队列调度关系的最后一个队列。如果不指定 **to** *end-queue-index* 可选参数，则仅为 *start-queue-index* 指定的队列配置队列调度关系，*start-queue-index* 与 *end-queue-index* 均为整数形式，取值范围是 0～7，队列优先级从 0 到 7 递增。 缺省情况下，队列对应的调度为 WFQ，可用 **undo schedule** 命令在队列模板中恢复各队列之间的调度关系为缺省配置
4	**quit** 例如：[Huawei-qos-queue-profile-profile1] **quit**	退出队列模板视图，返回系统视图

（续表）

步骤	命令	说明
5	**interface** *interface-type interface-number* 例如：[Huawei] **interface** ethernet 2/0/0	进入接口视图
6	**qos te queue-profile** *queue-profile-name* 例如：[Huawei-Ethernet2/0/0] **qos queue-profile** profile1	在接口下应用队列模板。参数 *queue-profile-name* 指定所应用的队列模板。 缺省情况下，接口上未应用队列模板，可用 **undo qos queue-profile** 命令删除在接口下应用的队列模板

9.2.7 DS-TE 隧道配置管理

已经完成静态或动态 DS-TE 隧道的所有配置后，可通过以下 **display** 命令查看相关配置，验证配置结果。

- **display mpls te ds-te** { **summary** | **te-class-mapping** [**default** | **config** | **verbose**] }：查看 DS-TE 相关信息。
- **display mpls te te-class-tunnel** { **all** | { **ct0** | **ct1** | **ct2** | **ct3** | **ct4** | **ct5** | **ct6** | **ct7** } **priority** *priority* }：查看 TE-CLASS 关联的 TE 隧道。
- **display interface tunnel** *interface-number*：查看隧道接口下各 CT 的流量信息。

9.2.8 Non-IETF 模式的 MAM 模型静态 DS-TE 配置示例

如图 9-8 所示，MPLS 骨干网的 PE 和 P 节点运行 OSPF 协议实现互通。PE1 和 PE2 接入 VPN-A 和 VPN-B。VPN-A 和 VPN-B 的流量分别为 EF 和 BE 类型。现要求在 PE1 和 PE2 之间建立 Non-IETF 模式的 DS-TE 静态 TE 隧道传输以上流量。要求各 CT 之间不允许发生带宽抢占。

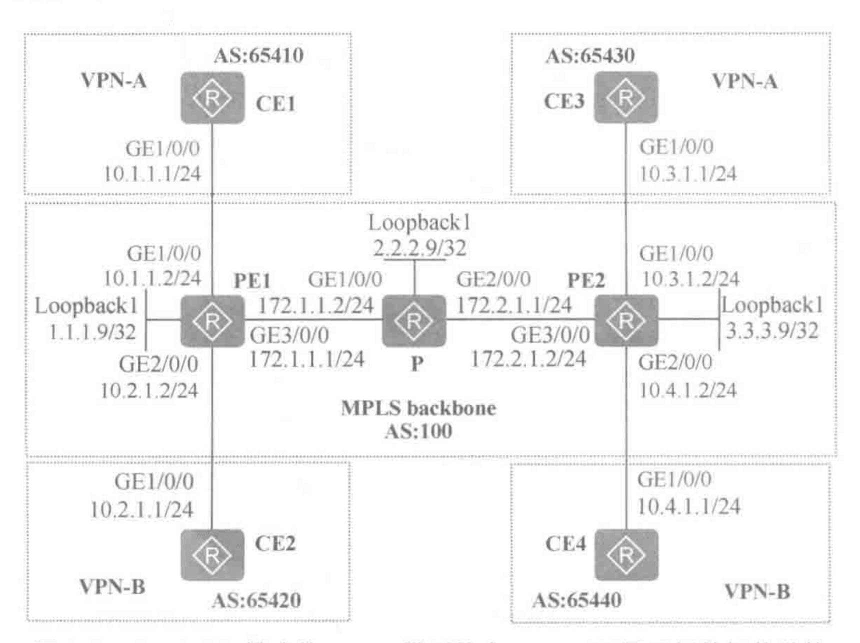


图 9-8 Non-IETF 模式的 MAM 模型静态 DS-TE 配置示例的拓扑结构

1. 基本配置思路分析

DS-TE 隧道是 TE 隧道的一种扩展，所以在建立 DS-TE 隧道之前也是需要完成 TE 隧道建立的基本配置，包括骨干网的公网 IGP 路由、全局和公网接口 MPLS 和 MPLS TE 能力的使能等。再结合 9.2.1 节所介绍的静态 DS-TE 隧道配置任务，可得出本示例的以下基本配置思路：

（1）在骨干网各节点上配置 Loopback 口和各公网接口 IP 地址，以及 OSPF 协议，实现骨干网三层互通。

（2）在骨干网各节点全局和公网接口上使能 MPLS、MPLS TE 能力。

（3）在两 PE 上创建 VPN 实例，名称分别为 VPN-A 和 VPN-B。PE1 上两 VPN 实例的 RD 分别为 100:1、100:2，双向 VPN-Target 属性值分别为 111:1、222:2；PE2 上两 VPN 实例的 RD 分别为 200:1、200:2，双向 VPN-Target 属性值分别为 111:1、222:2。

（4）在 PE 之间建立 MP-IBGP 对等体，在 PE 与 CE 之间建立 EBGP 对等体。

（5）在两 PE 上为两个 VPN 各自创建隧道接口，指定采用静态 CR-LSP 建立隧道。

（6）（可选）在两 PE 上配置隧道绑定策略，使所创建的 TE 隧道专用于 CT0 和 CT1 的业务传输。

说明 以上 6 个步骤都是用来建立 MPLS TE 隧道的，后面的各步是在 TE 隧道基础上配置 MPLS QoS 功能，完成 DS-TE 功能的实现。

（7）配置 DS-TE 工作模式为 Non-IETF，采用 MAM 带宽约束模型。

说明 因为本示例中的两个 VPN 实例中各只有一种业务类型，故 Non-IETF 模式就可满足要求，又要求 CT 间不发生抢占，故需要采用 MAM 带宽约束模型，使每种 CT 业务流量独占为自己的带宽。

（8）在 PE 和 P 节点上配置链路带宽。假设配置最大可预留总带宽为 375Mbit/s， 分配给 BC0 的带宽为 150Mbit/s、BC1 的带宽为 250Mbit/s。

（9）在骨干网各节点上建立 4 条（正、反方向各 2 条）静态 CR-LSP，在每条 CR-LSP 的 Ingress、Transit 节点上为 CT0 预留的带宽为 100Mbit/s，为 CT1 预留的带宽为 200Mbit/s。

（10）配置 CT 与业务类型的映射关系。在 PE 节点上配置入接口信任报文中的 DSCP 优先级，并将 DSCP 46（EF）映射到 LP0（对应 CT0），将 DSCP 0（BE）映射到 LP 1（对应 CT1），同时将 EXP 5 映射到 LP 0（对应 CT0），EXP 0 映射到 LP1（对应 CT1），实现 CT0 承载 EF 流量，CT1 承载 BE 流量。

说明 以上配置是假设 VPN-A 中流量的 DSCP 优先级值为 46，EXP 优先级值为 5；VPN-B 中流量的 DSCP 优先级值为 0，EXP 优先级值为 0。缺省情况下，CT0 所映射的 DSCP 值为 0～7、EXP 0，CT1 所映射的 DSCP 值为 8~15，EXP 1。因为实际所需的 DSCP、EXP 优先级与 CT 的映射关系与缺省映射关系不一致，故要重新配置。

（11）配置 CT 业务的调度方式，将 0 和 1 号队列（分别对应 CT0 和 CT1 业务类型）采用 WFQ 队列调度方式，其他队列采用 PQ 调度方式。其实本项配置任务也是可选的，

因为所有队列的缺省调度方式就是 WFQ。

2. 具体配置步骤

（1）在骨干网各节点上配置 Loopback 接口和公网接口（不包括连接 CE 的接口）的 IP 地址和 OSPF 路由，实现骨干网三层互通。

PE1 上的配置。

```
<Huawei> system-view
[Huawei] sysname PE1
[PE1] interface gigabitethernet 3/0/0
[PE1-GigabitEthernet3/0/0] ip address 172.1.1.1 255.255.255.0
[PE1-GigabitEthernet3/0/0] quit
[PE1] interface loopback 1
[PE1-LoopBack1] ip address 1.1.1.9 255.255.255.255
[PE1-LoopBack1] quit
[PE1] ospf 1
[PE1-ospf-1] area 0
[PE1-ospf-1-area-0.0.0.0] network 1.1.1.9 0.0.0.0
[PE1-ospf-1-area-0.0.0.0] network 172.1.1.0 0.0.0.255
[PE1-ospf-1-area-0.0.0.0] quit
[PE1-ospf-1] quit
```

P 上的配置。

```
<Huawei> system-view
[Huawei] sysname P
[P] interface gigabitethernet 1/0/0
[P-GigabitEthernet1/0/0] ip address 172.1.1.2 255.255.255.0
[P-GigabitEthernet1/0/0] quit
[P] interface gigabitethernet 2/0/0
[P-GigabitEthernet2/0/0] ip address 172.2.1.1 255.255.255.0
[P-GigabitEthernet2/0/0] quit
[P] interface loopback 1
[P-LoopBack1] ip address 2.2.2.9 255.255.255.255
[P-LoopBack1] quit
[P] ospf 1
[P-ospf-1] area 0
[P-ospf-1-area-0.0.0.0] network 2.2.2.9 0.0.0.0
[P-ospf-1-area-0.0.0.0] network 172.1.1.0 0.0.0.255
[P-ospf-1-area-0.0.0.0] network 172.2.1.0 0.0.0.255
[P-ospf-1-area-0.0.0.0] quit
[P-ospf-1] quit
```

PE2 上的配置。

```
<Huawei> system-view
[Huawei] sysname PE2
[PE2] interface gigabitethernet 3/0/0
[PE2-GigabitEthernet3/0/0] ip address 172.2.1.2 255.255.255.0
[PE2-GigabitEthernet3/0/0] quit
[PE2] interface loopback 1
[PE2-LoopBack1] ip address 3.3.3.9 255.255.255.255
[PE2-LoopBack1] quit
[PE2] ospf 1
[PE2-ospf-1] area 0
[PE2-ospf-1-area-0.0.0.0] network 3.3.3.9 0.0.0.0
[PE2-ospf-1-area-0.0.0.0] network 172.2.1.0 0.0.0.255
[PE2-ospf-1-area-0.0.0.0] quit
[PE2-ospf-1] quit
```

以上配置完成后，PE1、P、PE2 之间应能建立 OSPF 邻居关系，执行 **display ospf peer** 命令可以看到邻居状态为 Full。执行 **display ip routing-table** 命令可以看到 PE 之间学习到对方的 Loopback1 路由。

（2）在 PE 和 P 节点上配置 LSR-ID，并在全局和公网接口上使能 MPLS 和 MPLS TE 能力，建立 MPLS TE 公网隧道。

PE1 上的配置。

```
[PE1] mpls lsr-id 1.1.1.9
[PE1] mpls
[PE1-mpls] mpls te
[PE1-mpls] quit
[PE1] interface gigabitethernet 3/0/0
[PE1-GigabitEthernet3/0/0] mpls
[PE1-GigabitEthernet3/0/0] mpls te
[PE1-GigabitEthernet3/0/0] quit
```

P 上的配置。

```
[P] mpls lsr-id 2.2.2.9
[P] mpls
[P-mpls] mpls te
[P-mpls] quit
[P] interface gigabitethernet 1/0/0
[P-GigabitEthernet1/0/0] mpls
[P-GigabitEthernet1/0/0] mpls te
[P-GigabitEthernet1/0/0] quit
[P] interface gigabitethernet 2/0/0
[P-GigabitEthernet2/0/0] mpls
[P-GigabitEthernet2/0/0] mpls te
[P-GigabitEthernet2/0/0] quit
```

PE2 上的配置。

```
[PE2] mpls lsr-id 3.3.3.9
[PE2] mpls
[PE2-mpls] mpls te
[PE2-mpls] quit
[PE2] interface gigabitethernet 3/0/0
[PE2-GigabitEthernet3/0/0] mpls
[PE2-GigabitEthernet3/0/0] mpls te
[PE2-GigabitEthernet3/0/0] quit
```

（3）在 PE 上配置 VPN 实例，将 CE 接入 PE。

PE1 上的配置。配置两 VPN 实例的 RD 分别为 100:1 和 100:2，两 VPN 实例的 VPN-Target 属性分别为 111:1 和 222:2，并应用将在下一步配置的隧道绑定策略，使所创建的 TE 隧道只用于隧道绑定策略，将两个 VPN 实例绑定对应的 AC 接口。

```
[PE1] ip vpn-instance VPN-A
[PE1-vpn-instance-VPN-A] ipv4-family
[PE1-vpn-instance-VPN-A-af-ipv4] route-distinguisher 100:1
[PE1-vpn-instance-VPN-A-af-ipv4] vpn-target 111:1 both
[PE1-vpn-instance-VPN-A-af-ipv4] tnl-policy policya   #---配置采用名为 policya 的隧道策略
[PE1-vpn-instance-VPN-A-af-ipv4] quit
[PE1-vpn-instance-VPN-A] quit
[PE1] ip vpn-instance VPN-B
[PE1-vpn-instance-VPN-B] ipv4-family
[PE1-vpn-instance-VPN-B-af-ipv4] route-distinguisher 100:2
```

```
[PE1-vpn-instance-VPN-B-af-ipv4] vpn-target 222:2 both
[PE1-vpn-instance-VPN-B-af-ipv4] tnl-policy policyb
[PE1-vpn-instance-VPN-B-af-ipv4] quit
[PE1-vpn-instance-VPN-B] quit
[PE1] interface gigabitethernet 1/0/0
[PE1-GigabitEthernet1/0/0] ip binding vpn-instance VPN-A #---将 GE1/0/0 接口与 VPN-A 实例绑定
[PE1-GigabitEthernet1/0/0] ip address 10.1.1.2 24
[PE1-GigabitEthernet1/0/0] quit
[PE1] interface gigabitethernet 2/0/0
[PE1-GigabitEthernet2/0/0] ip binding vpn-instance VPN-B
[PE1-GigabitEthernet2/0/0] ip address 10.2.1.2 24
[PE1-GigabitEthernet2/0/0] quit
```

PE2 上的配置。配置两个 VPN 实例的 RD 分别为 200:1 和 200:2，两个 VPN 实例的 VPN-Target 属性分别为 111:1 和 222:2，并应用将在下一步配置的隧道绑定策略，使所创建的 TE 隧道只用于隧道绑定策略，将两个 VPN 实例绑定对应的 AC 接口。

```
[PE2] ip vpn-instance VPN-A
[PE2-vpn-instance-VPN-A] ipv4-family
[PE2-vpn-instance-VPN-A-af-ipv4] route-distinguisher 200:1
[PE2-vpn-instance-VPN-A-af-ipv4] vpn-target 111:1 both
[PE2-vpn-instance-VPN-A-af-ipv4] tnl-policy policya
[PE2-vpn-instance-VPN-A-af-ipv4] quit
[PE2-vpn-instance-VPN-A] quit
[PE2] ip vpn-instance VPN-B
[PE2-vpn-instance-VPN-B] ipv4-family
[PE2-vpn-instance-VPN-B-af-ipv4] route-distinguisher 200:2
[PE2-vpn-instance-VPN-B-af-ipv4] vpn-target 222:2 both
[PE2-vpn-instance-VPN-B-af-ipv4] tnl-policy policyb
[PE2-vpn-instance-VPN-B-af-ipv4] quit
[PE2-vpn-instance-VPN-B] quit
[PE2] interface gigabitethernet 1/0/0
[PE2-GigabitEthernet1/0/0] ip binding vpn-instance VPN-A
[PE2-GigabitEthernet1/0/0] ip address 10.3.1.2 24
[PE2-GigabitEthernet1/0/0] quit
[PE2] interface gigabitethernet 2/0/0
[PE2-GigabitEthernet2/0/0] ip binding vpn-instance VPN-B
[PE2-GigabitEthernet2/0/0] ip address 10.4.1.2 24
[PE2-GigabitEthernet2/0/0] quit
```

CE1 上的配置。

```
<Huawei> system-view
[Huawei] sysname CE1
[CE1] interface gigabitethernet 1/0/0
[CE1-GigabitEthernet1/0/0] ip address 10.1.1.1 255.255.255.0
[CE1-GigabitEthernet1/0/0] quit
```

CE2 上的配置。

```
<Huawei> system-view
[Huawei] sysname CE2
[CE2] interface gigabitethernet 1/0/0
[CE2-GigabitEthernet1/0/0] ip address 10.2.1.1 255.255.255.0
[CE2-GigabitEthernet1/0/0] quit
```

CE3 上的配置。

```
<Huawei> system-view
[Huawei] sysname CE3
```

```
[CE3] interface gigabitethernet 1/0/0
[CE3-GigabitEthernet1/0/0] ip address 10.3.1.1 255.255.255.0
[CE3-GigabitEthernet1/0/0] quit
```

CE4 上的配置。

```
<Huawei> system-view
[Huawei] sysname CE4
[CE4] interface gigabitethernet 1/0/0
[CE4-GigabitEthernet1/0/0] ip address 10.4.1.1 255.255.255.0
[CE4-GigabitEthernet1/0/0] quit
```

以上配置完成后，在 PE 上执行 **display ip vpn-instance verbose** 命令可以看到 VPN 实例的配置情况。

（4）在 PE 之间建立 MP-IBGP 对等体，PE 与 CE 之间建立 EBGP 对等体。

PE1 上的配置。

```
[PE1] bgp 100
[PE1-bgp] peer 3.3.3.9 as-number 100   #---与对等体 3.3.3.9（PE2）的 IBGP 对等体关系
[PE1-bgp] peer 3.3.3.9 connect-interface loopback 1   #---以 Loopback1 接口作为与对等体 3.3.3.9 进行 IBGP 会话的源接口
[PE1-bgp] ipv4-family vpnv4
[PE1-bgp-af-vpnv4] peer 3.3.3.9 enable   #---使能与对等体 3.3.3.9 的 VPN 路由信息交换能力
[PE1-bgp-af-vpnv4] quit
[PE1-bgp] ipv4-family vpn-instance VPN-A
[PE1-bgp-VPN-A] peer 10.1.1.1 as-number 65410   #---在 VPN-A 实例中与对等体 10.1.1.1（CE1）建立 EBGP 对等体关系
[PE1-bgp-VPN-A] import-route direct   #---引入直连路由
[PE1-bgp-VPN-A] quit
[PE1-bgp] ipv4-family vpn-instance VPN-B
[PE1-bgp-VPN-B] peer 10.2.1.1 as-number 65420
[PE1-bgp-VPN-B] import-route direct
[PE1-bgp-VPN-B] quit
```

PE2 上的配置。

```
[PE2] bgp 100
[PE2-bgp] peer 1.1.1.9 as-number 100
[PE2-bgp] peer 1.1.1.9 connect-interface loopback 1
[PE2-bgp] ipv4-family vpnv4
[PE2-bgp-af-vpnv4] peer 1.1.1.9 enable
[PE2-bgp-af-vpnv4] quit
[PE2-bgp] ipv4-family vpn-instance VPN-A
[PE2-bgp-VPN-A] peer 10.3.1.1 as-number 65430
[PE2-bgp-VPN-A] import-route direct
[PE2-bgp-VPN-A] quit
[PE2-bgp] ipv4-family vpn-instance VPN-B
[PE2-bgp-VPN-B] peer 10.4.1.1 as-number 65440
[PE2-bgp-VPN-B] import-route direct
[PE2-bgp-VPN-B] quit
```

CE1 上的配置。

```
[CE1] bgp 65410
[CE1-bgp] peer 10.1.1.2 as-number 100
[CE1-bgp] import-route direct
```

CE2 上的配置。

```
[CE2] bgp 65420
[CE2-bgp] peer 10.2.1.2 as-number 100
[CE2-bgp] import-route direct
```

CE3 上的配置。

```
[CE3] bgp 65430
[CE3-bgp] peer 10.3.1.2 as-number 100
[CE3-bgp] import-route direct
```

CE4 上的配置。

```
[CE4] bgp 65440
[CE4-bgp] peer 10.4.1.2 as-number 100
[CE4-bgp] import-route direct
```

以上配置完成后，在 PE 上执行 **display bgp vpnv4 all peer** 命令，可以看到 PE 之间、PE 与 CE 之间的 BGP 对等体关系已建立，并达到 Established 状态。以下是在 PE1 上执行该命令的输出示例（参见输出信息中的粗体字部分）。

```
[PE1] display bgp vpnv4 all peer
BGP local router ID : 1.1.1.9
 Local AS number : 100
 Total number of peers : 3                 Peers in established state : 3
  Peer         V    AS    MsgRcvd   MsgSent   OutQ   Up/Down       State        PrefRcv
  3.3.3.9      4    100   3         5         0      00:01:23      Established   0

  Peer of IPv4-family for vpn instance :

 VPN-Instance VPN-A, Router ID 1.1.1.9:
  10.1.1.1     4    65410 25        25        0      00:17:57    Established    1
 VPN-Instance VPN-B, Router ID 1.1.1.9:
  10.2.1.1     4    65420 21        22        0      00:17:10    Established    0
```

（5）在 PE1、PE2 上分别创建两个 Tunnel 接口，建立双向各两条 CR-LSP。

PE1 上的配置。

```
[PE1] interface tunnel 0/0/1
[PE1-Tunnel0/0/1] description For VPN-A_EF
[PE1-Tunnel0/0/1] ip address unnumbered interface loopback 1   #---指定 Tunnel0/0/1 接口借用 Loopback1 接口 IP 地址
[PE1-Tunnel0/0/1] tunnel-protocol mpls te   #---指定采用 MPLS TE 协议，建立 MPLS TE 隧道
[PE1-Tunnel0/0/1] destination 3.3.3.9
[PE1-Tunnel0/0/1] mpls te tunnel-id 300
[PE1-Tunnel0/0/1] mpls te signal-protocol cr-static   #---指定建立静态 CR-LSP
[PE1-Tunnel0/0/1] mpls te commit
[PE1-Tunnel0/0/1] quit
[PE1] interface tunnel 0/0/2
[PE1-Tunnel0/0/2] description For VPN-B_BE
[PE1-Tunnel0/0/2] ip address unnumbered interface loopback 1
[PE1-Tunnel0/0/2] tunnel-protocol mpls te
[PE1-Tunnel0/0/2] destination 3.3.3.9
[PE1-Tunnel0/0/2] mpls te tunnel-id 301
[PE1-Tunnel0/0/2] mpls te signal-protocol cr-static
[PE1-Tunnel0/0/2] mpls te commit
[PE1-Tunnel0/0/2] quit
```

PE2 上的配置。

```
[PE2] interface tunnel 0/0/1
[PE2-Tunnel0/0/1] description For VPN-A_EF
[PE2-Tunnel0/0/1] ip address unnumbered interface loopback 1
[PE2-Tunnel0/0/1] tunnel-protocol mpls te
[PE2-Tunnel0/0/1] destination 1.1.1.9
[PE2-Tunnel0/0/1] mpls te tunnel-id 300
[PE2-Tunnel0/0/1] mpls te signal-protocol cr-static
[PE2-Tunnel0/0/1] mpls te commit
```

```
[PE2-Tunnel0/0/1] quit
[PE2] interface tunnel 0/0/2
[PE2-Tunnel0/0/2] description For VPN-B_BE
[PE2-Tunnel0/0/2] ip address unnumbered interface loopback 1
[PE2-Tunnel0/0/2] tunnel-protocol mpls te
[PE2-Tunnel0/0/2] destination 1.1.1.9
[PE2-Tunnel0/0/2] mpls te tunnel-id 301
[PE2-Tunnel0/0/2] mpls te signal-protocol cr-static
[PE2-Tunnel0/0/2] mpls te commit
[PE2-Tunnel0/0/2] quit
```

（6）在两 PE 上配置隧道绑定策略，使指定的 Tunnel 接口与隧道目的 IP 地址进行绑定，从而限制该隧道只能承载特定的 VPN 业务。

PE1 上的配置。

```
[PE1] interface tunnel 0/0/1
[PE1-Tunnel0/0/1] mpls te reserved-for-binding   #---使能以上 Tunnel 接口对应的 TE 隧道只用于隧道绑定策略
[PE1-Tunnel0/0/1] mpls te commit
[PE1-Tunnel0/0/1] quit
[PE1] interface tunnel 0/0/2
[PE1-Tunnel0/0/2] mpls te reserved-for-binding
[PE1-Tunnel0/0/2] mpls te commit
[PE1-Tunnel0/0/2] quit
[PE1] tunnel-policy policya
[PE1-tunnel-policy-policya] tunnel binding destination 3.3.3.9 te tunnel 0/0/1   #---将 Tunnel0/0/1 接口与目的 IP 地址为
3.3.3.9 的 TE 隧道进行绑定
[PE1-tunnel-policy-policya] quit
[PE1] tunnel-policy policyb
[PE1-tunnel-policy-policyb] tunnel binding destination 3.3.3.9 te tunnel 0/0/2
[PE1-tunnel-policy-policyb] quit
```

PE2 上的配置。

```
[PE2] interface tunnel 0/0/1
[PE2-Tunnel0/0/1] mpls te reserved-for-binding
[PE2-Tunnel0/0/1] mpls te commit
[PE2-Tunnel0/0/1] quit
[PE2] interface tunnel 0/0/2
[PE2-Tunnel0/0/2] mpls te reserved-for-binding
[PE2-Tunnel0/0/2] mpls te commit
[PE2-Tunnel0/0/2] quit
[PE2] tunnel-policy policya
[PE2-tunnel-policy-policya] tunnel binding destination 1.1.1.9 te tunnel 0/0/1
[PE2-tunnel-policy-policya] quit
[PE2] tunnel-policy policyb
[PE2-tunnel-policy-policyb] tunnel binding destination 1.1.1.9 te tunnel 0/0/2
[PE2-tunnel-policy-policyb] quit
```

（7）在各节点上全局配置 DS-TE 模式和带宽约束模型，适用于设备上创建的所有 DS-TE 隧道。因为本示例仅需要对 EF 和 BE 这两种业务类型配置 QoS 服务，所以可以采用简单的 Non-IETF 模式，而本示例又要求各 CT 间不抢占带宽资源（静态 CR-LSP 也不支持带宽抢占），故要选择不支持抢占的 MAM 带宽约束模型。

PE1 上的配置。

```
[PE1] mpls
[PE1-mpls] mpls te ds-te mode non-ietf   #---配置 DS-TE 工作模式为 Non-IETF
[PE1-mpls] mpls te ds-te bcm mam   #---配置带宽约束模型为 MAM
```

```
[PE1-mpls] quit
```

P 上的配置。

```
[P] mpls
[P-mpls] mpls te ds-te mode non-ietf
[P-mpls] mpls te ds-te bcm mam
[P-mpls] quit
```

PE2 上的配置。

```
[PE2] mpls
[PE2-mpls] mpls te ds-te mode non-ietf
[PE2-mpls] mpls te ds-te bcm mam
[PE2-mpls] quit
```

以上配置完成后，在各节点上执行 **display mpls te ds-te summary** 命令，可查看 DS-TE 隧道的全局配置信息。以下是在 PE1 上执行该命令的输出示例。

```
[PE1] display mpls te ds-te summary
DS-TE IETF Supported :YES
DS-TE MODE                    :NON-IETF
Bandwidth Constraint Model    :MAM
```

（8）在各节点上配置隧道出方向链路的最大可用带宽，为 BC0、BC1 配置的预留带宽。在 MAM 带宽约束模型中，链路最大可预留带宽要大于或等于各 BC 的可预留带宽总和。现假设链路的最大可预留带宽均为 375Mbit/s，BC0、BC1 可预留的带宽分别为 125Mbit/s 和 250Mbit/s。本配置将同时适用于所有 CR-LSP。

PE1 上的配置。

```
[PE1] interface gigabitethernet 3/0/0
[PE1-GigabitEthernet3/0/0] mpls te bandwidth max-reservable-bandwidth 375000 #---配置链路可预留的总带宽为375000kbit/s
[PE1-GigabitEthernet3/0/0] mpls te bandwidth bc0 125000 bc1 250000 #---为 BC0 和 BC1 分别配置预留带宽为125000kbit/s 和 250000kit/s
[PE1-GigabitEthernet3/0/0] quit
```

P 上的配置。

```
[P] interface gigabitethernet 1/0/0
[P-GigabitEthernet1/0/0] mpls te bandwidth max-reservable-bandwidth 375000
[P-GigabitEthernet1/0/0] mpls te bandwidth bc0 125000 bc1 250000
[P-GigabitEthernet1/0/0] quit
[P] interface gigabitethernet 2/0/0
[P-GigabitEthernet2/0/0] mpls te bandwidth max-reservable-bandwidth 375000
[P-GigabitEthernet2/0/0] mpls te bandwidth bc0 125000 bc1 250000
[P-GigabitEthernet2/0/0] quit
```

PE2 上的配置。

```
[PE2] interface gigabitethernet 3/0/0
[PE2-GigabitEthernet3/0/0] mpls te bandwidth max-reservable-bandwidth 375000
[PE2-GigabitEthernet3/0/0] mpls te bandwidth bc0 125000 bc1 250000
[PE2-GigabitEthernet3/0/0] quit
```

以上配置完成后，在 PE 上执行 **display mpls te link-administration bandwidth-allocation** 命令，可查看链路的 BC 带宽分配情况。以下是在 PE1 上执行该命令的输出示例，但此时 CT0 和 CT1 上均没分配带宽，即 BW RESERVED（已保留的带宽）为 0（参见输出信息中的粗体字部分），因为此处并没有配置 CT0 和 CT1 的带宽。

在“BW AVAILABLE (kbit/sec)”字段中显示可以为 CT0 预留 125Mbit/s（等于 BC0 的带宽），为 CT1 预留 250Mbit/s（等于 BC1 的带宽），那是因为此时并没有实际建立

CR-LSP，缺省按一条 CR-LSP 来计算，即一个 BC 中就相当于只有一条 CR-LSP 中的对应 CT。

```
[PE1] display mpls te link-administration bandwidth-allocation interface gigabitethernet 3/0/0
  Link ID:   GigabitEthernet3/0/0
  Bandwidth Constraint Model     :  Maximum Allocation Model (MAM)
  Physical Link Bandwidth(Kbits/sec)                :  1000000
  Maximum Link Reservable Bandwidth(Kbits/sec):  375000
  Reservable Bandwidth BC0(kbits/sec)            :  125000
  Reservable Bandwidth BC1(kbits/sec)            :  250000
  Downstream Bandwidth (Kbits/sec)              :  0
  IPUpdown Link Status                          :  Up
  PhysicalUpdown Link Status                    :  Up
  GracefulUpdown Link Status                    :  DOWN
  ------------------------------------------------------------------------------
  TE-CLASS  CT  PRIORITY  BW RESERVED   BW AVAILABLE   DOWNSTREAM
                          (Kbit/sec)    (Kbit/sec)     RSVPLSPNODE COUNT
  ------------------------------------------------------------------------------
   0        0     0         0           125000           0
   1        0     1         0           125000           0
   2        0     2         0           125000           0
   3        0     3         0           125000           0
   4        0     4         0           125000           0
   5        0     5         0           125000           0
   6        0     6         0           125000           0
   7        0     7         0           125000           0
   8        1     0         0           250000           0
   9        1     1         0           250000           0
  10        1     2         0           250000           0
  11        1     3         0           250000           0
  12        1     4         0           250000           0
  13        1     5         0           250000           0
  14        1     6         0           250000           0
  15        1     7         0           250000           0
  ------------------------------------------------------------------------------
```

【经验提示】 *以上输出信息中只需关注 CT0 和 CT1 中的 0 优先级对应的预留带宽，因为静态 CR-LSP 的抢占优先级固定为 0。*

（9）在各节点上配置静态 CR-LSP，需要配置双向各 2 条静态 CR-LSP，分别以 PE1 和 PE2 为入节点。在 MAM/Extended-MAM 模型中，BC*i* 带宽值≥CT*i* 带宽值 x125%（0≤*i*≤7），在上一步已配置了 BC0 的可预留带宽为 125Mbit/s，BC1 的可预留带宽为 250Mbit/s，故可得出每条静态 CR-LSP 的 Ingress 和 Transit 节点的 CT0、CT1 的带宽分别为 100Mbit/s 和 200Mbit/s。

PE1 上的配置。PE1 对于两条相反方向的 CR-LSP 分别担当 Ingress 和 Egress 节点，要分别建立对应的 CR-LSP。

```
[PE1] static-cr-lsp ingress tunnel-interface tunnel 0/0/1 destination 3.3.3.9 nexthop 172.1.1.2 out-label 100 bandwidth ct0 100000
[PE1] static-cr-lsp ingress tunnel-interface tunnel 0/0/2 destination 3.3.3.9 nexthop 172.1.1.2 out-label 200 bandwidth ct1 200000
[PE1] static-cr-lsp egress VPN-A_EF incoming-interface gigabitethernet 3/0/0 in-label 101
[PE1] static-cr-lsp egress VPN-B_BE incoming-interface gigabitethernet 3/0/0 in-label 201
```

P 上的配置。

```
[P] static-cr-lsp transit VPN-A_EF-1to2 incoming-interface gigabitethernet1/0/0 in-label 100 nexthop 172.2.1.2 out-label 100 bandwidth ct0 100000
[P] static-cr-lsp transit VPN-B_BE-1to2 incoming-interface gigabitethernet1/0/0 in-label 200 nexthop 172.2.1.2 out-label 200 bandwidth ct1 200000
[P] static-cr-lsp transit VPN-A_EF-2to1 incoming-interface gigabitethernet2/0/0 in-label 101 nexthop 172.1.1.1 out-label 101 bandwidth ct0 100000
[P] static-cr-lsp transit VPN-B_BE-2to1 incoming-interface gigabitethernet2/0/0 in-label 201 nexthop 172.1.1.1 out-label 201 bandwidth ct1 200000
```

PE2 上的配置。PE2 对于两条相反方向的 CR-LSP 分别担当 Ingress 和 Egress 节点，要分别建立对应的 CR-LSP。

```
[PE2] static-cr-lsp egress VPN-A_EF incoming-interface gigabitethernet 3/0/0 in-label 100
[PE2] static-cr-lsp egress VPN-B_BE incoming-interface gigabitethernet 3/0/0 in-label 200
[PE2] static-cr-lsp ingress tunnel-interface tunnel 0/0/1 destination 1.1.1.9 nexthop 172.2.1.1 out-label 101 bandwidth ct0 100000
[PE2] static-cr-lsp ingress tunnel-interface tunnel 0/0/2 destination 1.1.1.9 nexthop 172.2.1.1 out-label 201 bandwidth ct1 200000
```

以上配置步骤完成后，在 PE 上执行 **display mpls static-cr-lsp** 命令，可发现前面所创建的静态 CR-LSP 的状态均为 Up。以下是在 PE1 上执行该命令查看 Tunnel0/0/1 接口上创建的静态 CR-LSP 状态的输出示例。

```
[PE1] display mpls static-cr-lsp Tunnel0/0/1
TOTAL          : 1      STATIC CRLSP(S)
Up             : 1      STATIC CRLSP(S)
DOWN           : 0      STATIC CRLSP(S)
Name           FEC              I/O Label     I/O If              Status
Tunnel0/0/1    3.3.3.9/32       NULL/100      -/GE3/0/0           Up
```

此时，再次执行 **display mpls te link-administration bandwidth-allocation** 命令，查看链路的带宽分配情况，可发现为优先级为 0 的 CT0 和 CT1 分配了带宽。

```
[PE1] display mpls te link-administration bandwidth-allocation interface gigabitethernet 3/0/0
  Link ID:  GigabitEthernet3/0/0
  Bandwidth Constraint Model    :  Maximum Allocation Model (MAM)
  Physical Link Bandwidth(kbits/sec)              :  1000000
  Maximum Link Reservable Bandwidth(kbits/sec):  375000
  Reservable Bandwidth BC0(kbits/sec)             :  125000
  Reservable Bandwidth BC1(kbits/sec)             :  250000
  Downstream Bandwidth (kbits/sec)                :  300000
  IPUpdown Link Status                            :  Up
  PhysicalUpdown Link Status                      :  Up
  GracefulUpdown Link Status                      :  DOWN
  ------------------------------------------------------------------------------
  TE-CLASS  CT  PRIORITY  BW RESERVED   BW AVAILABLE   DOWNSTREAM
                          (kbit/sec)    (kbit/sec)   RSVPLSPNODE COUNT
  ------------------------------------------------------------------------------
     0      0      0       100000         25000           0
     1      0      1       0              25000           0
     2      0      2       0              25000           0
     3      0      3       0              25000           0
     4      0      4       0              25000           0
     5      0      5       0              25000           0
     6      0      6       0              25000           0
     7      0      7       0              25000           0
     8      1      0       200000         50000           0
     9      1      1       0              50000           0
```

```
  10    1     2        0         50000         0
  11    1     3        0         50000         0
  12    1     4        0         50000         0
  13    1     5        0         50000         0
  14    1     6        0         50000         0
  15    1     7        0         50000         0
 ----------------------------------------------------------------------------------------
```

【经验提示】对比第（8）步配置完成后执行 **display mpls te link-administration bandwidth-allocation** 命令的输出结果可以看出，本步配置完成后再执行该命令的输出有以下两方面的变化。

■ CT0和CT1中的"BW RESERVED"字段值由原来的0变成了实际配置的值——100Mbit/s、200Mbit/s。那是因为第（8）步还没建立CR-LSP，配置的是BC0和BC1的带宽，并没有配置CT0和CT1的带宽，而本步是真正针对每一条CR-LSP配置了CT带宽。

■ CT0和CT1中的"BW AVAILABLE"字段值是原来的2倍，即分别由原来的125Mbit/s、250Mbit/s变成了250Mbit/s和500Mbit/s。那是因为在第（8）步的配置仅是针对一条CR-LSP的配置。而本步完成后在PE1上作为Ingress节的CR-LSP有两条，每条250Mbit/s，所以最终就是500Mbit/s了。

（10）在PE1、PE2节点上配置各入接口信任的报文优先级（Ingress和Egress节点信任报文中的DSCP优先级，Transit节点信任报文中的EXP优先级），并修改VPN-A和VPN-B中的DSCP 46（EF）、DSCP 0（BE）分别与LP 0、LP 1优先级进行映射，EXP 5、EXP 0分别与LP 0、LP 1优先级进行映射。

PE1上的配置。

```
[PE1] interface gigabitethernet 1/0/0
[PE1-GigabitEthernet1/0/0] trust dscp    #---信任IP报文中的DSCP优先级
[PE1-GigabitEthernet1/0/0] quit
[PE1] interface gigabitethernet 2/0/0
[PE1-GigabitEthernet2/0/0] trust dscp
[PE1-GigabitEthernet2/0/0] quit
[PE1] interface gigabitethernet 3/0/0
[PE1-GigabitEthernet3/0/0] trust dscp
[PE1-GigabitEthernet3/0/0] quit
[PE1] qos map-table dscp-lp
[PE1-maptbl-dscp-lp] input 46 output 0   #---将DSCP 46映射为LP 0
[PE1-maptbl-dscp-lp] input 0 output 1    #---将DSCP 0映射为LP 1
[PE1-maptbl-dscp-lp] quit
[PE1] qos map-table exp-lp
[PE1-maptbl-exp-lp] input 5 output 0
[PE1-maptbl-exp-lp] input 0 output 1
[PE1-maptbl-exp-lp] quit
```

P上的配置。

```
[P] interface gigabitethernet 1/0/0
[P-GigabitEthernet1/0/0] trust exp
[P-GigabitEthernet1/0/0] quit
[P] interface gigabitethernet 2/0/0
[P-GigabitEthernet2/0/0] trust exp
[P-GigabitEthernet2/0/0] quit
[P] qos map-table dscp-lp
```

```
[P-maptbl-dscp-lp] input 46 output 0
[P-maptbl-dscp-lp] input 0 output 1
[P-maptbl-dscp-lp] quit
[P] qos map-table exp-lp
[P-maptbl-exp-lp] input 5 output 0
[P-maptbl-exp-lp] input 0 output 1
[P-maptbl-exp-lp] quit
```

PE2 上的配置。

```
[PE2] interface gigabitethernet 1/0/0
[PE2-GigabitEthernet1/0/0] trust dscp
[PE2-GigabitEthernet1/0/0] quit
[PE2] interface gigabitethernet 2/0/0
[PE2-GigabitEthernet2/0/0] trust dscp
[PE2-GigabitEthernet2/0/0] quit
[PE2] interface gigabitethernet 3/0/0
[PE2-GigabitEthernet3/0/0] trust dscp
[PE2-GigabitEthernet3/0/0] quit
[PE2] qos map-table dscp-lp
[PE2-maptbl-dscp-lp] input 46 output 0
[PE2-maptbl-dscp-lp] input 0 output 1
[PE2-maptbl-dscp-lp] quit
[PE2] qos map-table exp-lp
[PE2-maptbl-exp-lp] input 5 output 0
[PE2-maptbl-exp-lp] input 0 output 1
[PE2-maptbl-exp-lp] quit
```

在 PE 上执行 **display qos map-table dscp-lp** 命令，可查看 DSCP 到本地优先级（LP）的映射关系。以下是在 PE1 上执行该命令的输出示例，对比 9.1.6 节表 9-3 中 DSCP 与 LP 的缺省映射关系可以看出，只有 DSCP 0 和 DSCP 46 与 LP 的映射关系发生了变化（参见输出信息中的粗体字部分）。

```
[PE1] display qos map-table dscp-lp
Input DSCP        LP
-------------------
0                 1
1                 0
2                 0
3                 0
4                 0
5                 0
6                 0
7                 0
8                 1
...
46                0
47                5
48                6
49                6
50                6
51                6
52                6
53                6
54                6
55                6
56                7
```

```
57          7
58          7
59          7
60          7
61          7
62          7
63          7
```

在 PE 上执行 **display qos map-table exp-lp** 命令，可查看 EXP 到本地优先级（LP）的映射关系。以下是在 PE1 上执行该命令的输出示例，与 9.1.6 节表 9-3 中的 EXP 与 LP 缺省映射关系相比，发生变化的也只有 EXP 0 和 EXP 5（参见输出信息中的粗体字部分）。

```
[PE1] display qos map-table exp-lp
Input EXP      LP
-------------------
 0             1
 1             1
 2             2
 3             3
 4             4
 5             0
 6             6
 7             7
```

（11）配置 CT 业务的调度方式。

本示例中 VPN-A 中的流量对应 LP0，VPN-B 中的流量对应 LP1，缺省均采用 WFQ 调度方式。现要求 VPN-A 和 VPN-B 进入的队列 0、1 的流量均采用 WFQ 调度方式，其他队列采用 PQ 调度方式。然后在骨干网各节点的出接口上应用。注意，因为通信是双向的，所以 P 节点的两个接口都可以是出接口，都需要配置。

PE1 上的配置。

```
[PE1] qos queue-profile queue-profile1
[PE1-qos-queue-profile-queue-profile1] schedule wfq 0 to 1 pq 2 to 7
[PE1-qos-queue-profile-queue-profile1] quit
[PE1] interface gigabitethernet 3/0/0
[PE1-GigabitEthernet3/0/0] qos te queue-profile queue-profile1
[PE1-GigabitEthernet3/0/0] quit
```

P 上的配置。

```
[P] qos queue-profile queue-profile1
[P-qos-queue-profile-queue-profile1] schedule wfq 0 to 1 pq 2 to 7
[P-qos-queue-profile-queue-profile1] quit
[P] interface gigabitethernet 1/0/0
[P-GigabitEthernet1/0/0] qos te queue-profile queue-profile1
[P-GigabitEthernet1/0/0] quit
[P] interface gigabitethernet 2/0/0
[P-GigabitEthernet2/0/0] qos te queue-profile queue-profile1
[P-GigabitEthernet2/0/0] quit
```

PE2 上的配置。

```
[PE2] qos queue-profile queue-profile1
[PE2-qos-queue-profile-queue-profile1] schedule wfq 0 to 1 pq 2 to 7
[PE2-qos-queue-profile-queue-profile1] quit
[PE2] interface gigabitethernet 3/0/0
[PE2-GigabitEthernet3/0/0] qos te queue-profile queue-profile1
[PE2-GigabitEthernet3/0/0] quit
```

以上配置步骤完成后，在 PE 上执行 **display qos queue-profile** 命令，可查看已配置的队列模板信息，以下是在 PE1 上执行该命令的输出示例。

```
[PE1] display qos queue-profile queue-profile1
Queue-profile: queue-profile1
Queue   Schedule   Weight   Length(Bytes/Packets) GTS(CIR/CBS)
---------------------------------------------------------------------------------
0       WFQ        10            -/-             -/-
1       WFQ        10            -/-             -/-
2       PQ         -             -/-             -/-
3       PQ         -             -/-             -/-
4       PQ         -             -/-             -/-
5       PQ         -             -/-             -/-
6       PQ         -             -/-             -/-
7       PQ         -             -/-             -/-
```

9.3 动态 DS-TE 隧道配置与管理

上节介绍的静态 DS-TE 与第 5 章介绍的静态 TE 一样，CR-LSP 的建立都是采用手动静态配置的，无需信令协议。本节介绍的动态 DS-TE 与第 5 章介绍的动态 TE 一样，都是使用 RSVP-TE 作为 CR-LSP 建立的信令协议。动态 DS-TE 隧道可以根据网络变化动态改变，在规模较大的组网中，可以避免逐跳配置的麻烦。

动态 DS-TE 也是在动态 TE 的基础上进行的，所以其配置任务和配置方法与第 5 章介绍的动态 TE 基本一样，只是增多了一些 MPLS QoS 方面的功能配置。而在许多 QoS 功能配置方面又与上节介绍的静态 DS-TE 中的 QoS 功能方面的配置方法完全一样，可直接参考。动态 DS-TE 所包括的配置任务具体如下。

（1）使能 MPLS TE 和 RSVP-TE。与第 5 章 5.7.1 节介绍的配置方法完全一样，参见即可。

（2）配置 MPLS TE 隧道接口。与第 5 章 5.7.2 节介绍的配置方法完全一样，参见即可。

（3）配置 DS-TE 模式和带宽约束模型。与 9.2.2 节介绍的配置方法完全一样，参见即可。

（4）（可选）配置 TE-Class 映射表和链路带宽。与 9.2.3 节介绍的配置方法完全一样，参见即可。

（5）配置 TE 信息发布。与第 5 章 5.7.3 节介绍的配置方法完全一样，参见即可。

（6）配置动态 DS-TE 隧道的约束条件。

（7）配置路径计算。与第 5 章 5.7.5 节介绍的配置方法完全一样，参见即可。

（8）配置接口信任的报文优先级。与 9.2.5 节介绍的配置方法完全一样，参见即可。

（9）（可选）配置 CT 与业务类型的映射关系以及调度方式。与 9.2.6 节介绍的配置方法完全一样，参见即可。

下面仅介绍上述第（6）项配置任务的具体配置方法。

配置动态 DS-TE 隧道之前，需要完成以下任务。

■ 配置 IGP 路由协议，使各节点间的 IP 路由可达。
■ 配置各 LSR 节点的 LSR-ID。
■ 配置各 LSR 节点的全局 MPLS 能力。
■ 配置各 LSR 节点的接口 MPLS 能力。

9.3.1　配置动态 DS-TE 隧道的约束条件

在约束条件方面，动态 DS-TE 隧道与 TE 隧道非常类似，都是在隧道入节点配置显式路径和各 CT 约束带宽这两方面。但 DS-TE 隧道与 TE 隧道相比，支持更多的 CT，所以在 CT 约束带宽配置方面有所不同。

动态 DS-TE 隧道的显式路径和 CT 约束带宽的具体配置步骤如表 9-13 所示。配置各 CT 约束带宽时，要求所有 CR-LSP 中的同类 CT 的约束带宽总和不超过对应的 BC 的带宽，即 CT*i* 只能使用 BC*i* 的带宽。

表 9-13　　动态 DS-TE 隧道约束条件的配置步骤

步骤	命令	说明
1	**system-view** 例如：<Huawei> **system-view**	进入系统视图
2	**explicit-path** *path-name* 例如：[Huawei] **explicit-path** p1	创建显式路径，进入显式路径视图。参数 *path-name* 用来指定隧道的显式路径名称，字符串形式，不区分大小写，不支持空格，长度范围为大于等于 1。 【注意】*必须启动 MPLS TE 功能后才能配置隧道的显式路径。且显式路径上的节点地址不能重复，也不能形成环路。如果有环路，CSPF 将检测出环路，无法成功计算出路径。* 缺省情况下，没有配置隧道的显式路径，可用 **undo explicit-path** *path-name* 命令删除配置的指定显式路径
3	**next hop** *ip-address* [**include** [[**loose** | **strict**] | [**incoming** | **outgoing**]]* | **exclude**] 例如：[Huawei-explicit-path-p1] **next hop** 10.0.0.125 **exclude**	指定显式路径的下一个节点。命令中的参数和选项说明如下。 ● *ip-address*：指定显式路径中的下一个节点 IP 地址。 ● **include** [[**loose** | **strict**] | [**incoming** | **outgoing**]]*：二选一可选项，指定在显式路径中包含此节点。其中的各可多选选项如下。 　● **strict**：表示严格显式路径，参数 *ip-address* 指定的节点与本节点必须直连。缺省情况下，采用 **strict** 模式，即加入的下一跳与上一节点必须是直连的。作为一种约束条件，显式路径可以指定经过或者不经过某些节点。 　● **loose**：表示松散显式路径，参数 *ip-address* 指定的节点与本节点可以不是直连的。 　● **incoming**：指定参数 *ip-address* 为当前配置的下一个节点的入接口地址。 　● **outgoing**：指定参数 *ip-address* 为当前配置的下一个节点的出接口地址。 ● **exclude**：二选一可选项，指定显式路径不能经过参数 *ip-address* 指定的节点。

（续表）

步骤	命令	说明
3	**next hop** *ip-address* [**include** [[**loose** | **strict**] | [**incoming** | **outgoing**]] * | **exclude**] 例如：[Huawei-explicit-path-p1] **next hop** 10.0.0.125 **exclude**	【说明】需通过本命令依次把路径中的每个下一跳列出来，构建完整的显式路径。如果指定的 *ip-address* 是当前配置的下一个节点的入接口地址，建议配置 **incoming** 参数；如果指定的 *ip-address* 是当前配置的下一个节点的出接口地址，建议配置 **outgoing** 参数。 缺省情况下，没有在显式路径中指定下一个节点，可用 **undo next hop** *ip-address* 命令删除指定的下一跳
执行以下命令增加、修改或删除显式路径中的节点		
4	**list hop** [*ip-address*] 例如：[Huawei-explicit-path-path1] **list hop**	查看显式路径节点信息。可选参数 *ip-address* 用来指定要查看当前显式路径配置的节点的 IP 地址。如果不指定本参数，则查看当前显式路径下的所有节点
	add hop *ip-address1* [**include** [[**loose** | **strict**] | [**incoming** | **outgoing**]] * | **exclude**] { **after** | **before** } *ip-address2* 例如：[Huawei-explicit-path-p1] **add hop** 10.2.2.2 **exclude after** 10.1.1.1	（可选）向显式路径中插入一个节点。本命令中的大多数参数和选项与第 3 步中的命令中的参数和选项一样，只不过这里是插入节点的操作，下面仅介绍不同的参数和选项。 • **after**：二选一选项，表示在参数 *ip-address2* 后插入参数 *ip-address1* 指定的节点。 • **before**：二选一选项，表示在参数 *ip-address2* 前插入参数 *ip-address1* 指定的节点。 • *ip-address2*：指定已经在显式路径中的节点接口 IP 地址或节点 Router ID。 如果指定的 *ip-address1* 是新增节点的入接口地址，建议配置 **incoming** 参数；如果指定的 *ip-address1* 是新增节点的出接口地址，建议配置 **outgoing** 参数
	modify hop *ip-address1 ip-address2* [**include** [[**loose** | **strict**] | [**incoming** | **outgoing**]] * | exclude] 例如：[Huawei-explicit-path-p1] **modify hop** 1.1.1.9 2.2.2.9	（可选）修改显式路径中的节点地址。参数 *ip-address1 ip-address2* 指定将显式路径中的 IP 地址 *ip-address1* 修改为 *ip-address2*。其他选项与本表第 3 步中的对应选项作用一样。 【说明】如果指定的 *ip-address2* 是修改后节点的入接口地址，建议配置 **incoming** 参数；如果指定的 *ip-address2* 是修改后节点的出接口地址，建议配置 **outgoing** 参数
	delete hop *ip-address* 例如：[Huawei-explicit-path-p1] **delete hop** 10.10.10.10	（可选）从显式路径中删除一个节点。参数 *ip-address* 用来指定要删除节点的 IP 地址。此节点必须是显式路径中存在的节点
5	**quit** 例如：[Huawei-explicit-path-p1] **quit**	返回系统视图
6	**interface tunnel** *tunnel-number* 例如：[Huawei] **interface tunnel** 0/0/1	进入 MPLS TE 隧道的 Tunnel 接口视图

（续表）

步骤	命令	说明
7	**mpls te bandwidth** { **ct0** *ct0-bw-value* | **ct1** *ct1-bw-value* } 例如：[Huawei-Tunnel0/0/1] **mpls te bandwidth ct0** 2000	（多选一）在非标准（Non-IETF）模式单 CT 情形下配置 Tunnel 接口带宽。命令中的参数说明如下。 ● **ct0** *ct0-bw-value*：二选一参数，指定为 CT0 预留带宽，整数形式，取值范围是 1～4000000000，单位是 kbit/s。 ● **ct1** *ct1-bw-value*：二选一参数，指定为 CT1 预留带宽，整数形式，取值范围是 1～4000000000，单位是 kbit/s。 缺省情况下，没有配置隧道带宽，可用 **undo mpls te bandwidth** { **all** | **ct0** [*ct0-bw-value*] | **ct1** [*ct1-bw-value*] } 命令恢复指定 CT 或所有 CT 的预留带宽为缺省设置
	mpls te bandwidth { **ct0** *bw-value* | **ct1** *bw-value* | **ct2** *bw-value* | **ct3** *bw-value* | **ct4** *bw-value* | **ct5** *bw-value* | **ct6** *bw-value* | **ct7** *bw-value* } 例如：[Huawei-Tunnel0/0/1] **mpls te bandwidth ct5** 2000	（多选一）在标准（IETF）模式多 CT 情形下配置 Tunnel 接口带宽。每个的参数分别是为 CT0~CT7（只能选择其中一个）设置预留带宽，整数形式，取值范围都是 1～4000000000，单位是 kbit/s 缺省情况下，没有配置隧道带宽，可用 **undo mpls te bandwidth** { **all** | { **ct0** *ct0-bw-value* | **ct1** *ct1-bw-value* | **ct2** *ct2-bw-value* | **ct3** *ct3-bw-value* | **ct4** *ct4-bw-value* | **ct5** *ct5-bw-value* | **ct6** *ct6-bw-value* | **ct7** *ct7-bw-value* } }命令恢复指定 CT 或所有 CT 的预留带宽为缺省设置
	mpls te bandwidth { **ct0** *bw-value* | **ct1** *bw-value* | **ct2** *bw-value* | **ct3** *bw-value* | **ct4** *bw-value* | **ct5** *bw-value* | **ct6** *bw-value* | **ct7** *bw-value* } * 例如：[Huawei-Tunnel0/0/1] **mpls te bandwidth ct0** 1000 **ct5** 2000	（多选一）在标准（IETF）模式多 CT 情形下配置 Tunnel 接口带宽。每个的参数分别是为 CT0~CT7（可选择多个）设置预留带宽，整数形式，取值范围都是 1～4000000000，单位是 kbit/s。 缺省情况下，没有配置隧道带宽，可用 **undo mpls te bandwidth** { **all** | { **ct0** *ct0-bw-value* | **ct1** *ct1-bw-value* | **ct2** *ct2-bw-value* | **ct3** *ct3-bw-value* | **ct4** *ct4-bw-value* | **ct5** *ct5-bw-value* | **ct6** *ct6-bw-value* | **ct7** *ct7-bw-value* } *}命令恢复指定 CT 或所有 CT 的预留带宽为缺省设置
8	**mpls te path explicit-path** *path-name* 例如：[Huawei-Tunnel0/0/1] **mpls te path explicit-path** p1	配置隧道应用的显式路径，该路径是在本表第 2 步创建的显式路径。 缺省情况下，没有为当前隧道配置显式路径，**undo mpls te path explicit-path** *path-name* 命令用来删除显式路径
9	**mpls te commit** 例如：[Huawei-Tunnel0/0/1] **mpls te commit**	提交隧道当前配置，使以上配置生效

说明　在隧道策略中，多 CT 的 CR-LSP 只支持隧道绑定策略，不支持隧道选择策略。同一个节点，无论使用哪种带宽约束模型，所有 CTi 的带宽总和不超过 BCi 的带宽值（$0 \leq i \leq 7$），即 CTi 只能使用 BCi 的带宽。例如，某 PE 节点的 BC1 带宽值为 x，该节点一共有两条 CT1 的 CR-LSP，带宽分别为 y 和 z，则 $y+z \leq x$。

9.3.2　RDM 模型 IETF 模式的动态 DS-TE 配置示例

如图 9-9 所示，MPLS 骨干网的 PE 和 P 节点运行 OSPF 协议实现互通，P 节点不支持 MPLS LDP。PE1 和 PE2 接入 VPN-A 和 VPN-B。PE3 和 PE4 之间流量通过普通的 TE 隧道承载。要求在 PE1 和 PE2 之间建立 DS-TE 隧道传递以上流量，并满足以下各类型

流量的 QoS 需求。

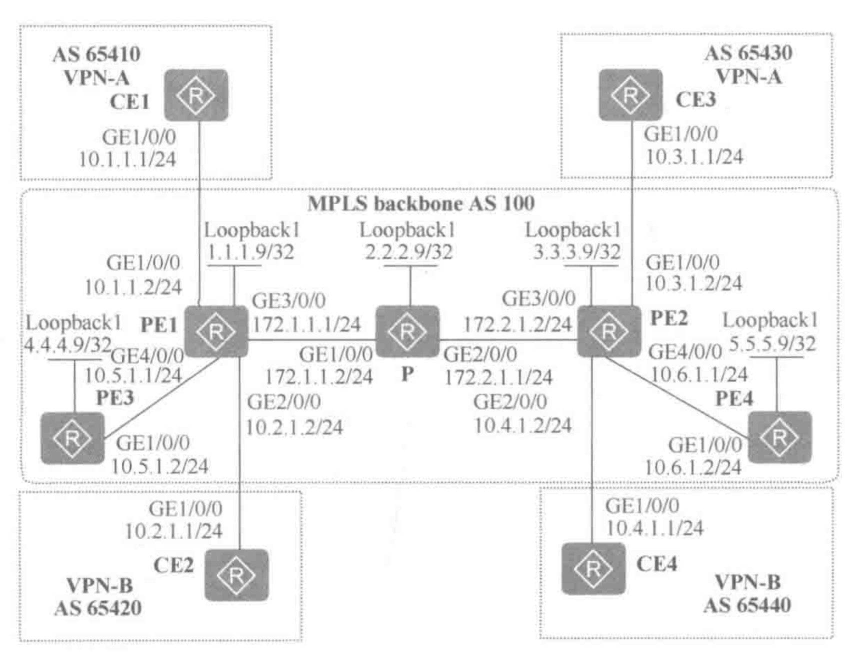


图 9-9 RDM 模型 IETF 模式的动态 DS-TE 配置示例的拓扑结构

VPN-A 中的流量有 AF2、AF1 两者类型；VPN-B 中的流量有 AF2、AF1、BE 三种类型；PE3 与 PE4 之间的流量为 BE 类型。各类型流量的 QoS 需求如表 9-14 所示，即 BE 流的 DSCP 优先级值为 DSCP 0，AF1 流的 DSCP 优先级值为 DSCP 10，AF2 流的 DSCP 优先级值为 DSCP 20。带宽约束模型要求为 RDM，允许 CTi 抢占低优先级的 CTj 的带宽（$0 \leqslant i<j \leqslant 7$），以确保高优先级 CT 的带宽。

表 9-14　　示例中各类型流量的 QoS 需求

流类型	带宽	抖动
VPN-A 的 AF2 流（DSCP20）	100Mbit/s	小于 50ms
VPN-A 的 AF1 流（DSCP10）	50Mbit/s	小于 200ms
VPN-B 的 AF2 流（DSCP20）	100Mbit/s	小于 50ms
VPN-B 的 AF1 流（DSCP10）	50Mbit/s	小于 200ms
VPN-B 的 BE 流（DSCP0）	50Mbit/s	无需求
PE3 与 PE4 之间的 BE 流（DSCP0）	50Mbit/s	无需求

1. 基本配置思路分析

本示例中，VPN-A 和 VPN-B 的流量类型部分相同（都有 AF2 流和 AF1 流），所以需要用两条 TE 隧道分别承载 VPN-A 和 VPN-B 中的流量。VPN-B 和 PE3 与 PE4 之间的流量类型也有部分相同（都有 BE 流），因此也需要用不同隧道承载。但 VPN-A 和 PE3 与 PE4 之间的流量类型各不相同，故可用同一条 TE 隧道承载。这样一来，在 P1 和 PE2 上只需要创建两条 TE 隧道，一条用来承载 VPN-A 和 PE3 与 PE4 之间的流量，包括 AF1、AF2 和 BE 三种流量，一条用于承载 VPN-B 中的流量，也包括 AF1、AF2、BE 这三种流量。

本示例只需在 PE1、P 和 PE2 之间建立 DS-TE 隧道，在 PE3 和 PE1 之间，以及 PE4 和 PE2 之间仅需建立普通的 LDP LSP 隧道即可。

根据以上分析，再结合 9.2.1 节介绍的动态 DS-TE 隧道配置任务，可得出本示例的基本配置思路如下。

（1）在各 PE 和 P 节点上配置各接口（包括 Loopback 接口）的 IP 地址，并通过 OSPF 协议实现各 PE 和 P 节点的三层互通。

（2）在各 PE 和 P 节点上配置 LSR-ID、使能 MPLS，并在 PE1、PE2 和 P 上使能 MPLS TE 和 RSVP-TE，在各 PE 上使能 MPLS LDP，以使在 PE1、P 和 PE2 节点间建立 MPLS TE 隧道，在 PE3 和 PE1 之间、PE4 和 PE2 之间建立 LDP LSP 隧道。

（3）在 PE1、PE2 之间建立 MP-IBGP 对等体关系，使 PE1 和 PE2 之间可直接交互 BGP Update 消息；在 PE1、PE2 与各 CE 之间建立 EBGP 对等体关系。

（4）在 PE1 和 PE2 间创建两个 DS-TE 隧道接口：Tunnel0/0/1 和 Tunnel0/0/2。其中，每个隧道配置三个 CT，分别为 CT0、CT1 和 CT2，对应的优先级都为 0。假设 CT0、CT1 和 CT2 的可预留带宽分别为 50Mbit/s、50Mbit/s 和 100Mbit/s，分别用于承载 BE、AF1 和 AF2 流。

VPN-A 的 AF2、AF1 流分别使用 Tunnel0/0/1 的 CT2、CT1 承载。PE3 与 PE4 之间的 BE 流使用 Tunnel0/0/1 的 CT0 承载。VPN-B 的 AF2、AF1、BE 流分别使用 Tunnel0/0/2 的 CT2、CT1 和 CT0 承载。

（5）在 PE1、PE2 和 P 上配置 OSPF TE，并使能 CSPF，发布 TE 信息，计算 TE 隧道路径。

（6）在 PE1、PE2 上创建两个 VPN 实例，并配置相关属性，使对应的 CE 站点加入到指定的 VPN 实例中。

（7）在 PE1、PE2 上配置隧道策略，使到达指定目的 IP 地址的 TE 隧道仅可用于传输本示例中指定的业务。

（8）在 PE1、PE2 和 P 节点上配置 DS-TE 模式和带宽约束模型。因为本示例中每个 VPN 中有三种 CT 业务类型，故只能采用标准的 IETF 模式。本示例中又要求各 CT 间可共享带宽，故可采用 RDM 带宽约束模型。

（9）在 PE1、PE2 和 P 节点上配置链路带宽。本示例采用 RDM 带宽约束模型，又由于两条隧道的路径一样，因此链路上 BCi 带宽应不小于所有 TE 隧道的 CT*i*～CT7 带宽的总和，且链路的最大可预留带宽应不小于 BC0 带宽。

在第（4）步已分别为 CT0、CT1 和 CT2 配置了 50Mbit/s、50Mbit/s 和 100Mbit/s 的可预留带宽。根据 10.3.3 节介绍的 RDM 约束带宽模型精确控制流量带宽计算方法 BC*i* 带宽值≥CT*i*～CT7 的总带宽值 × 125%（$0 \leqslant i \leqslant 7$），可得出各 BC 的带宽值。

- BC2 的带宽≥125% ×（Tunnel0/0/1 的 CT2＋Tunnel0/0/2 的 CT2）＝250Mbit/s。
- BC1 的带宽≥BC2 的带宽＋125% ×（Tunnel0/0/1 的 CT1＋Tunnel0/0/2 的 CT1）＝375Mbit/s。
- BC0 的带宽≥BC1 的带宽＋125% ×（Tunnel0/0/1 的 CT0＋Tunnel0/0/2 的 CT0）＝500Mbit/s。
- 链路可预留带宽≥BC0 的带宽＝500Mbit/s。

（10）在 PE 上配置 TE-Class 映射表。

TE-Class 映射表是全局概念，TE-Class 映射表应用到该 LSR 的所有 DS-TE 隧道中。

在本示例的两条隧道中只有 CT0、CT1 和 CT2 三种业务类型，所以可只配置它们对应的 TE-Class0、TE-Class1 和 TE-Class2 映射，因为与缺省 TE-Class 映射表配置一样（参见第 9.1.3 节表 9-1），故可不进行本项配置任务。

（11）在 PE1 和 PE2 上配置显式路径，指定两条 CP-LSP 所经过节点。

（12）在 PE1、PE2 和 P 节点上配置入接口信任的报文优先级以及优先级映射。PE 节点信息 DSCP 优先级，P 节点信任 EXP 优先级。从表 9-21 可以看出，本示例中的三种类型流量对应的 DSCP 优先级、LP 优先级与 EXP 优先级之间的映射关系恰好与 9.1.6 节中表 9-3 所示的缺省映射关系一致，故不用配置。

（13）配置 CT 业务的调度方式。本示例假设 对 CT0 中的 AF1 流量和 CT1 中的 AF2 流量采取 WFQ 调度方式，对 CT2 中的 BE 流量采用 PQ 调度方式，以满足 AF1 和 AF2 精武业务流量对时延抖动的更高要求。

（14）在 TE 隧道入节点上配置转发邻接，将 CR-LSP 发布给邻居节点，其他节点能够使用此隧道，引入流量到对应的隧道中。

2. 具体配置步骤

（1）在各 PE 和 P 节点上配置各接口 IP 地址，并配置 OSPF，实现骨干网的三层互通。

PE1 上的配置。

```
<Huawei> system-view
[Huawei] sysname PE1
[PE1] interface gigabitethernet 3/0/0
[PE1-GigabitEthernet3/0/0] ip address 172.1.1.1 255.255.255.0
[PE1-GigabitEthernet3/0/0] quit
[PE1] interface gigabitethernet 4/0/0
[PE1-GigabitEthernet4/0/0] ip address 10.5.1.1 255.255.255.0
[PE1-GigabitEthernet4/0/0] quit
[PE1] interface loopback 1
[PE1-LoopBack1] ip address 1.1.1.9 255.255.255.255
[PE1-LoopBack1] quit
[PE1] ospf 1
[PE1-ospf-1] area 0
[PE1-ospf-1-area-0.0.0.0] network 1.1.1.9 0.0.0.0
[PE1-ospf-1-area-0.0.0.0] network 172.1.1.0 0.0.0.255
[PE1-ospf-1-area-0.0.0.0] network 10.5.1.0 0.0.0.255
[PE1-ospf-1-area-0.0.0.0] quit
[PE1-ospf-1] quit
```

P 上的配置。

```
<Huawei> system-view
[Huawei] sysname P
[P] interface gigabitethernet 1/0/0
[P-GigabitEthernet1/0/0] ip address 172.1.1.2 255.255.255.0
[P-GigabitEthernet1/0/0] quit
[P] interface gigabitethernet 2/0/0
[P-GigabitEthernet2/0/0] ip address 172.2.1.2 255.255.255.0
[P-GigabitEthernet2/0/0] quit
[P] interface loopback 1
[P-LoopBack1] ip address 2.2.2.9 255.255.255.255
[P-LoopBack1] quit
[P] ospf 1
[P-ospf-1] area 0
```

```
[P-ospf-1-area-0.0.0.0] network 2.2.2.9 0.0.0.0
[P-ospf-1-area-0.0.0.0] network 172.1.1.0 0.0.0.255
[P-ospf-1-area-0.0.0.0] network 172.2.1.0 0.0.0.255
[P-ospf-1-area-0.0.0.0] quit
[P-ospf-1] quit
```

PE2 上的配置。

```
<Huawei> system-view
[Huawei] sysname PE2
[PE2] interface gigabitethernet 3/0/0
[PE2-GigabitEthernet3/0/0] ip address 172.2.1.2 255.255.255.0
[PE2-GigabitEthernet3/0/0] quit
[PE2] interface gigabitethernet 4/0/0
[PE2-GigabitEthernet4/0/0] ip address 10.6.1.1 255.255.255.0
[PE2-GigabitEthernet4/0/0] quit
[PE2] interface loopback 1
[PE2-LoopBack1] ip address 3.3.3.9 255.255.255.255
[PE2-LoopBack1] quit
[PE2] ospf 1
[PE2-ospf-1] area 0
[PE2-ospf-1-area-0.0.0.0] network 3.3.3.9 0.0.0.0
[PE2-ospf-1-area-0.0.0.0] network 172.2.1.0 0.0.0.255
[PE2-ospf-1-area-0.0.0.0] network 10.6.1.0 0.0.0.255
[PE2-ospf-1-area-0.0.0.0] quit
[PE2-ospf-1] quit
```

PE3 上的配置。

```
<Huawei> system-view
[Huawei] sysname PE3
[PE3] interface gigabitethernet 1/0/0
[PE3-GigabitEthernet1/0/0] ip address 10.5.1.2 255.255.255.0
[PE3-GigabitEthernet1/0/0] quit
[PE3] interface loopback 1
[PE3-LoopBack1] ip address 4.4.4.9 255.255.255.255
[PE3-LoopBack1] quit
[PE3] ospf 1
[PE3-ospf-1] area 0
[PE3-ospf-1-area-0.0.0.0] network 4.4.4.9 0.0.0.0
[PE3-ospf-1-area-0.0.0.0] network 10.5.1.0 0.0.0.255
[PE3-ospf-1-area-0.0.0.0] quit
[PE3-ospf-1] quit
```

PE4 上的配置。

```
<Huawei> system-view
[Huawei] sysname PE4
[PE4] interface gigabitethernet 1/0/0
[PE4-GigabitEthernet1/0/0] ip address 10.6.1.2 255.255.255.0
[PE4-GigabitEthernet1/0/0] quit
[PE4] interface loopback 1
[PE4-LoopBack1] ip address 5.5.5.9 255.255.255.255
[PE4-LoopBack1] quit
[PE4] ospf 1
[PE4-ospf-1] area 0
[PE4-ospf-1-area-0.0.0.0] network 5.5.5.9 0.0.0.0
[PE4-ospf-1-area-0.0.0.0] network 10.6.1.0 0.0.0.255
[PE4-ospf-1-area-0.0.0.0] quit
[PE4-ospf-1] quit
```

以上配置完成后，各节点之间应能建立 OSPF 邻居关系，执行 **display ospf peer** 命令可以看到邻居状态为 Full。执行 **display ip routing-table** 命令可以看到 PE 之间学习到对方的 Loopback1 路由。

（2）在所有 PE 和 P 节点上配置 LSR-ID、使能 MPLS，并在 PE1、PE2 和 P 上使能 MPLS TE 和 RSVP-TE（用于通过 RSVP-TE 建立 CR-LSP），在所有 PE 上使能 MPLS LDP，以使 PE 间建立 LDP LSP（但在 P 节点上不用使能 LDP，因为 PE1 和 PE2 之间是通过 RSVP-TE 协议建立 CR-LSP 的）。

PE1 上的配置。

```
[PE1] mpls lsr-id 1.1.1.9
[PE1] mpls
[PE1-mpls] mpls te
[PE1-mpls] mpls rsvp-te
[PE1-mpls] quit
[PE1] mpls ldp
[PE1-mpls-ldp] quit
[PE1] interface gigabitethernet 3/0/0
[PE1-GigabitEthernet3/0/0] mpls
[PE1-GigabitEthernet3/0/0] mpls te
[PE1-GigabitEthernet3/0/0] mpls rsvp-te
[PE1-GigabitEthernet3/0/0] quit
[PE1] interface gigabitethernet 4/0/0
[PE1-GigabitEthernet4/0/0] mpls
[PE1-GigabitEthernet4/0/0] mpls ldp
[PE1-GigabitEthernet4/0/0] quit
```

P 上的配置。

```
[P] mpls lsr-id 2.2.2.9
[P] mpls
[P-mpls] mpls te
[P-mpls] mpls rsvp-te
[P-mpls] quit
[P] interface gigabitethernet 1/0/0
[P-GigabitEthernet1/0/0] mpls
[P-GigabitEthernet1/0/0] mpls te
[P-GigabitEthernet1/0/0] mpls rsvp-te
[P-GigabitEthernet1/0/0] quit
[P] interface gigabitethernet 2/0/0
[P-GigabitEthernet2/0/0] mpls
[P-GigabitEthernet2/0/0] mpls te
[P-GigabitEthernet2/0/0] mpls rsvp-te
[P-GigabitEthernet2/0/0] quit
```

PE2 上的配置。

```
[PE2] mpls lsr-id 3.3.3.9
[PE2] mpls
[PE2-mpls] mpls te
[PE2-mpls] mpls rsvp-te
[PE2-mpls] quit
[PE2] mpls ldp
[PE2-mpls-ldp] quit
[PE2] interface gigabitethernet 3/0/0
[PE2-GigabitEthernet3/0/0] mpls
[PE2-GigabitEthernet3/0/0] mpls te
```

```
[PE2-GigabitEthernet3/0/0] mpls rsvp-te
[PE2-GigabitEthernet3/0/0] quit
[PE2] interface gigabitethernet 4/0/0
[PE2-GigabitEthernet4/0/0] mpls
[PE2-GigabitEthernet4/0/0] mpls ldp
[PE2-GigabitEthernet4/0/0] quit
```

PE3 上的配置。

```
[PE3] mpls lsr-id 4.4.4.9
[PE3] mpls
[PE3-mpls] quit
[PE3] mpls ldp
[PE3-mpls-ldp] quit
[PE3] interface gigabitethernet 1/0/0
[PE3-GigabitEthernet1/0/0] mpls
[PE3-GigabitEthernet1/0/0] mpls ldp
[PE3-GigabitEthernet1/0/0] quit
```

PE4 上的配置。

```
[PE4] mpls lsr-id 5.5.5.9
[PE4] mpls
[PE4-mpls] quit
[PE4] mpls ldp
[PE4-mpls-ldp] quit
[PE4] interface gigabitethernet 1/0/0
[PE4-GigabitEthernet1/0/0] mpls
[PE4-GigabitEthernet1/0/0] mpls ldp
[PE4-GigabitEthernet1/0/0] quit
```

以上配置完成后，在 PE1、PE2 或 P 节点上执行 **display mpls rsvp-te interface** 命令，可查看使能了 RSVP 的接口及 RSVP 相关信息。在 PE1、PE2、PE3 或 PE4 上执行命令 **display mpls ldp lsp**，可发现 PE3 和 PE1 之间，以及 PE2 和 PE4 之间均存在一条 LDP LSP。

（3）在 PE1 与 PE2 之间建立 MP-IBGP 对等体，在 PE1、PE2 与各 CE 之间建立 EBGP 对等体。

PE1 上的配置。

```
[PE1] bgp 100
[PE1-bgp] peer 3.3.3.9 as-number 100
[PE1-bgp] peer 3.3.3.9 connect-interface loopback 1
[PE1-bgp] ipv4-family vpnv4
[PE1-bgp-af-vpnv4] peer 3.3.3.9 enable
[PE1-bgp-af-vpnv4] quit
[PE1-bgp] ipv4-family vpn-instance VPN-A
[PE1-bgp-VPN-A] peer 10.1.1.1 as-number 65410
[PE1-bgp-VPN-A] import-route direct
[PE1-bgp-VPN-A] quit
[PE1-bgp] ipv4-family vpn-instance VPN-B
[PE1-bgp-VPN-B] peer 10.2.1.1 as-number 65420
[PE1-bgp-VPN-B] import-route direct
[PE1-bgp-VPN-B] quit
```

PE2 上的配置。

```
[PE2] bgp 100
[PE2-bgp] peer 1.1.1.9 as-number 100
[PE2-bgp] peer 1.1.1.9 connect-interface loopback 1
[PE2-bgp] ipv4-family vpnv4
```

```
[PE2-bgp-af-vpnv4] peer 1.1.1.9 enable
[PE2-bgp-af-vpnv4] quit
[PE2-bgp] ipv4-family vpn-instance VPN-A
[PE2-bgp-VPN-A] peer 10.3.1.1 as-number 65430
[PE2-bgp-VPN-A] import-route direct
[PE2-bgp-VPN-A] quit
[PE2-bgp] ipv4-family vpn-instance VPN-B
[PE2-bgp-VPN-B] peer 10.4.1.1 as-number 65440
[PE2-bgp-VPN-B] import-route direct
[PE2-bgp-VPN-B] quit
```

CE1 上的配置。

```
[CE1] bgp 65410
[CE1-bgp] peer 10.1.1.2 as-number 100
[CE1-bgp] import-route direct
```

CE2 上的配置。

```
[CE2] bgp 65420
[CE2-bgp] peer 10.2.1.2 as-number 100
[CE2-bgp] import-route direct
```

CE3 上的配置。

```
[CE3] bgp 65430
[CE3-bgp] peer 10.3.1.2 as-number 100
[CE3-bgp] import-route direct
```

CE4 上的配置。

```
[CE4] bgp 65440
[CE4-bgp] peer 10.4.1.2 as-number 100
[CE4-bgp] import-route direct
```

以上配置完成后，在 PE 上执行 **display bgp vpnv4 all peer** 命令，可以看到 PE 之间的 BGP 对等体关系已建立，并达到 Established 状态。以下是在 PE1 上执行该命令的输出示例。

```
[PE1] display bgp vpnv4 all peer
BGP local router ID : 1.1.1.9
 Local AS number : 100
 Total number of peers : 3                    Peers in established state : 3
  Peer          V    AS    MsgRcvd   MsgSent   OutQ   Up/Down        State PrefRcv
  3.3.3.9       4    100   3          5          0    00:01:23    Established    0

  Peer of IPv4-family for vpn instance :

 VPN-Instance VPN-A, Router ID 1.1.1.9:
  10.1.1.1      4    65410 25         25         0    00:17:57    Established    1
 VPN-Instance VPN-B, Router ID 1.1.1.9:
  10.2.1.1      4    65420 21         22         0    00:17:10    Established    0
```

（4）在 PE1、PE2 上各创建两个 Tunnel 接口，分别建立用于传输 VPN-A，以及 PE3 和 PE4 之间流量的 TE 隧道，以及用于传输 VPN-B 中流量的 TE 隧道。同进配置两条隧道中的 CT0、CT1 和 CT2 的带宽分别为 50Mbit/s、50Mbit/s 和 100Mbit/s。

PE1 上的配置。

```
[PE1] interface tunnel 0/0/1
[PE1-Tunnel0/0/1] description For VPN-A & Non-VPN
[PE1-Tunnel0/0/1] ip address unnumbered interface loopback 1
[PE1-Tunnel0/0/1] tunnel-protocol mpls te
```

```
[PE1-Tunnel0/0/1] destination 3.3.3.9
[PE1-Tunnel0/0/1] mpls te tunnel-id 300
[PE1-Tunnel0/0/1] mpls te signal-protocol rsvp-te   #---指定采用 RSVP-TE 作为信令协议建立 CR-LSP
[PE1-Tunnel0/0/1] mpls te path explicit-path path1   #---指定采用 path1 中指定的显式路径
[PE1-Tunnel0/0/1] mpls te priority 0 0   #---指定隧道的建立优先级和保持优先级均为 0
[PE1-Tunnel0/0/1] mpls te bandwidth ct0 50000 ct1 50000 ct2 100000
[PE1-Tunnel0/0/1] mpls te commit
[PE1-Tunnel0/0/1] quit
[PE1] interface tunnel 0/0/2
[PE1-Tunnel0/0/2] description For VPN-B
[PE1-Tunnel0/0/2] ip address unnumbered interface loopback 1
[PE1-Tunnel0/0/2] tunnel-protocol mpls te
[PE1-Tunnel0/0/2] destination 3.3.3.9
[PE1-Tunnel0/0/2] mpls te tunnel-id 301
[PE1-Tunnel0/0/2] mpls te signal-protocol rsvp-te
[PE1-Tunnel0/0/2] mpls te path explicit-path path1
[PE1-Tunnel0/0/2] mpls te priority 0 0
[PE1-Tunnel0/0/2] mpls te bandwidth ct0 50000 ct1 50000 ct2 100000
[PE1-Tunnel0/0/2] mpls te commit
[PE1-Tunnel0/0/2] quit
```

PE2 上的配置。

```
[PE2] interface tunnel 0/0/1
[PE2-Tunnel0/0/1] description For VPN-A & Non-VPN
[PE2-Tunnel0/0/1] ip address unnumbered interface loopback 1
[PE2-Tunnel0/0/1] tunnel-protocol mpls te
[PE2-Tunnel0/0/1] destination 1.1.1.9
[PE2-Tunnel0/0/1] mpls te tunnel-id 300
[PE2-Tunnel0/0/1] mpls te signal-protocol rsvp-te
[PE2-Tunnel0/0/1] mpls te path explicit-path path1
[PE2-Tunnel0/0/1] mpls te priority 0 0
[PE2-Tunnel0/0/1] mpls te bandwidth ct0 50000 ct1 50000 ct2 100000
[PE2-Tunnel0/0/1] mpls te commit
[PE2-Tunnel0/0/1] quit
[PE2] interface tunnel 0/0/2
[PE2-Tunnel0/0/2] description For VPN-B
[PE2-Tunnel0/0/2] ip address unnumbered interface loopback 1
[PE2-Tunnel0/0/2] tunnel-protocol mpls te
[PE2-Tunnel0/0/2] destination 1.1.1.9
[PE2-Tunnel0/0/2] mpls te tunnel-id 301
[PE2-Tunnel0/0/2] mpls te signal-protocol rsvp-te
[PE2-Tunnel0/0/2] mpls te path explicit-path path1
[PE2-Tunnel0/0/2] mpls te priority 0 0
[PE2-Tunnel0/0/2] mpls te bandwidth ct0 50000 ct1 50000 ct2 100000
[PE2-Tunnel0/0/2] mpls te commit
[PE2-Tunnel0/0/2] quit
```

以上配置完成后，在 PE 上执行 **display interface tunnel** *interface-number* 命令，可发现 Tunnel 接口为 Up 状态。以下是在 PE1 上执行该命令输出示例。

```
[PE1] display interface tunnel 0/0/1
Tunnel0/0/1 current state : Up
Line protocol current state : Up
Last line protocol up time : 2013-01-06 20:24:46
Description:For VPN-A & Non-VPN
Route Port,The Maximum Transmit Unit is 1500
Internet Address is unnumbered, using address of LoopBack1(1.1.1.9/32)
```

```
Encapsulation is TUNNEL, loopback not set
Tunnel destination 3.3.3.9
Tunnel up/down statistics 1
Tunnel protocol/transport MPLS/MPLS, ILM is available,
primary tunnel id is 0x6, secondary tunnel id is 0x0
Current system time: 2013-01-06 20:29:02
    300 seconds output rate 0 bits/sec, 0 packets/sec
    0 seconds output rate 0 bits/sec, 0 packets/sec
    0 packets output,  0 bytes
    0 output error
    0 output drop
    ct0:0 packets output,  0 bytes
        0 output error
       0 packets output drop
    ct1:0 packets output,  0 bytes
        0 output error
       0 packets output drop
    ct2:0 packets output,  0 bytes
        0 output error
       0 packets output drop
    Input bandwidth utilization  : 0%
    Output bandwidth utilization : 0%
```

在 PE 上执行 **display mpls te te-class-tunnel** 命令，可查看 TE-CLASS 关联的 TE 隧道。以下是在 PE1 上执行该命令的输出示例。

```
[PE1] display mpls te te-class-tunnel all
---------------------------------------------------------------------------
No.        CT  priority  status        tunnel name          tunnel commit
---------------------------------------------------------------------------
1          0   0           Valid          Tunnel0/0/1          Yes
2          0   0           Valid          Tunnel0/0/2          Yes
3          1   0           Valid          Tunnel0/0/1          Yes
4          1   0           Valid          Tunnel0/0/2          Yes
5          2   0           Valid          Tunnel0/0/1          Yes
6          2   0           Valid          Tunnel0/0/2          Yes
```

（5）在 PE1、PE2 和 P 上配置 OSPF TE，并在入节点上使能 CSPF，以通过 OSPF TE 发布私网路由信息。

PE1 上的配置。

```
[PE1] ospf 1
[PE1-ospf-1] opaque-capability enable
[PE1-ospf-1] area 0
[PE1-ospf-1-area-0.0.0.0] mpls-te enable
[PE1-ospf-1-area-0.0.0.0] quit
[PE1-ospf-1] quit
[PE1] mpls
[PE1-mpls] mpls te cspf
```

P 上的配置。

```
[P] ospf 1
[P-ospf-1] opaque-capability enable
[P-ospf-1] area 0
[P-ospf-1-area-0.0.0.0] mpls-te enable
[P-ospf-1-area-0.0.0.0] quit
[P-ospf-1] quit
```

PE2上的配置。

```
[PE2] ospf 1
[PE2-ospf-1] opaque-capability enable
[PE2-ospf-1] area 0
[PE2-ospf-1-area-0.0.0.0] mpls-te enable
[PE2-ospf-1-area-0.0.0.0] quit
[PE2-ospf-1] quit
[PE2] mpls
[PE2-mpls] mpls te cspf
[PE2-mpls] quit
```

以上配置完成后，在PE或P节点上执行 **display ospf mpls-te** 命令，可查看OSPF链路状态数据库中包含的TE LSA信息。

（6）在PE1和PE2上配置VPN实例，将CE接入PE。

PE1上的配置。配置两VPN实例的RD分别为100:1和100:2，两VPN实例的VPN-Target属性分别为111:1和222:2，并应用将在下一步配置的隧道绑定策略，使所创建的TE隧道只用于隧道绑定策略，将两VPN实例绑定对应的AC接口。

```
[PE1] ip vpn-instance VPN-A
[PE1-vpn-instance-VPN-A] ipv4-family
[PE1-vpn-instance-VPN-A-af-ipv4] route-distinguisher 100:1
[PE1-vpn-instance-VPN-A-af-ipv4] vpn-target 111:1 both
[PE1-vpn-instance-VPN-A-af-ipv4] tnl-policy policya
[PE1-vpn-instance-VPN-A-af-ipv4] quit
[PE1-vpn-instance-VPN-A] quit
[PE1] ip vpn-instance VPN-B
[PE1-vpn-instance-VPN-B] ipv4-family
[PE1-vpn-instance-VPN-B-af-ipv4] route-distinguisher 100:2
[PE1-vpn-instance-VPN-B-af-ipv4] vpn-target 222:2 both
[PE1-vpn-instance-VPN-B-af-ipv4] tnl-policy policyb
[PE1-vpn-instance-VPN-B-af-ipv4] quit
[PE1-vpn-instance-VPN-B] quit
[PE1] interface gigabitethernet 1/0/0
[PE1-GigabitEthernet1/0/0] ip binding vpn-instance VPN-A
[PE1-GigabitEthernet1/0/0] ip address 10.1.1.2 24
[PE1-GigabitEthernet1/0/0] quit
[PE1] interface gigabitethernet 2/0/0
[PE1-GigabitEthernet2/0/0] ip binding vpn-instance VPN-B
[PE1-GigabitEthernet2/0/0] ip address 10.2.1.2 24
[PE1-GigabitEthernet2/0/0] quit
```

PE2上的配置。配置两个VPN实例的RD分别为200:1和200:2，两个VPN实例的VPN-Target属性分别为111:1和222:2，并应用将在下一步配置的隧道绑定策略，使所创建的TE隧道只用于隧道绑定策略，将两个VPN实例绑定对应的AC接口。

```
[PE2] ip vpn-instance VPN-A
[PE2-vpn-instance-VPN-A] ipv4-family
[PE2-vpn-instance-VPN-A-af-ipv4] route-distinguisher 200:1
[PE2-vpn-instance-VPN-A-af-ipv4] vpn-target 111:1 both
[PE2-vpn-instance-VPN-A-af-ipv4] tnl-policy policya
[PE2-vpn-instance-VPN-A-af-ipv4] quit
[PE2-vpn-instance-VPN-A] quit
[PE2] ip vpn-instance VPN-B
[PE2-vpn-instance-VPN-B] ipv4-family
[PE2-vpn-instance-VPN-B-af-ipv4] route-distinguisher 200:2
```

```
[PE2-vpn-instance-VPN-B-af-ipv4] vpn-target 222:2 both
[PE2-vpn-instance-VPN-B-af-ipv4] tnl-policy policyb
[PE2-vpn-instance-VPN-B-af-ipv4] quit
[PE2-vpn-instance-VPN-B] quit
[PE2] interface gigabitethernet 1/0/0
[PE2-GigabitEthernet1/0/0] ip binding vpn-instance VPN-A
[PE2-GigabitEthernet1/0/0] ip address 10.3.1.2 24
[PE2-GigabitEthernet1/0/0] quit
[PE2] interface gigabitethernet 2/0/0
[PE2-GigabitEthernet2/0/0] ip binding vpn-instance VPN-B
[PE2-GigabitEthernet2/0/0] ip address 10.4.1.2 24
[PE2-GigabitEthernet2/0/0] quit
```

CE1 上的配置。

```
<Huawei> system-view
[Huawei] sysname CE1
[CE1] interface gigabitethernet 1/0/0
[CE1-GigabitEthernet1/0/0] ip address 10.1.1.1 255.255.255.0
[CE1-GigabitEthernet1/0/0] quit
```

CE2 上的配置。

```
<Huawei> system-view
[Huawei] sysname CE2
[CE2] interface gigabitethernet 1/0/0
[CE2-GigabitEthernet1/0/0] ip address 10.2.1.1 255.255.255.0
[CE2-GigabitEthernet1/0/0] quit
```

CE2 上的配置。

```
<Huawei> system-view
[Huawei] sysname CE3
[CE3] interface gigabitethernet 1/0/0
[CE3-GigabitEthernet1/0/0] ip address 10.3.1.1 255.255.255.0
[CE3-GigabitEthernet1/0/0] quit
```

CE4 上的配置。

```
<Huawei> system-view
[Huawei] sysname CE4
[CE4] interface gigabitethernet 1/0/0
[CE4-GigabitEthernet1/0/0] ip address 10.4.1.1 255.255.255.0
[CE4-GigabitEthernet1/0/0] quit
```

以上配置完成后，在 PE 上执行 **display ip vpn-instance verbose** 命令可以看到 VPN 实例的配置情况。

（7）在 PE1 和 PE2 上配置隧道绑定策略，使到达指定目的地址的 TE 隧道仅用于传输特定流量。

PE1 上的配置。

```
[PE1] interface tunnel 0/0/1
[PE1-Tunnel0/0/1] mpls te reserved-for-binding
[PE1-Tunnel0/0/1] mpls te commit
[PE1-Tunnel0/0/1] quit
[PE1] interface tunnel 0/0/2
[PE1-Tunnel0/0/2] mpls te reserved-for-binding
[PE1-Tunnel0/0/2] mpls te commit
[PE1-Tunnel0/0/2] quit
[PE1] tunnel-policy policya
[PE1-tunnel-policy-policya] tunnel binding destination 3.3.3.9 te tunnel 0/0/1
```

```
[PE1-tunnel-policy-policya] quit
[PE1] tunnel-policy policyb
[PE1-tunnel-policy-policyb] tunnel binding destination 3.3.3.9 te tunnel 0/0/2
[PE1-tunnel-policy-policyb] quit
```

PE2 上的配置。

```
[PE2] interface tunnel 0/0/1
[PE2-Tunnel0/0/1] mpls te reserved-for-binding
[PE2-Tunnel0/0/1] mpls te commit
[PE2-Tunnel0/0/1] quit
[PE2] interface tunnel 0/0/2
[PE2-Tunnel0/0/2] mpls te reserved-for-binding
[PE2-Tunnel0/0/2] mpls te commit
[PE2-Tunnel0/0/2] quit
[PE2] tunnel-policy policya
[PE2-tunnel-policy-policya] tunnel binding destination 1.1.1.9 te tunnel 0/0/1
[PE2-tunnel-policy-policya] quit
[PE2] tunnel-policy policyb
[PE2-tunnel-policy-policyb] tunnel binding destination 1.1.1.9 te tunnel 0/0/2
[PE2-tunnel-policy-policyb] quit
```

（8）在 PE1、PE2 和 P 节点上配置 IETF DS-TE 隧道模式和 RDM 带宽约束模型。

PE1 上的配置。

```
[PE1] mpls
[PE1-mpls] mpls te ds-te mode ietf
[PE1-mpls] mpls te ds-te bcm rdm
[PE1-mpls] quit
```

P 上的配置。

```
[P] mpls
[P-mpls] mpls te ds-te mode ietf
[P-mpls] mpls te ds-te bcm rdm
[P-mpls] quit
```

PE2 上的配置。

```
[PE2] mpls
[PE2-mpls] mpls te ds-te mode ietf
[PE2-mpls] mpls te ds-te bcm rdm
[PE2-mpls] quit
```

以上配置完成后，在 PE 或 P 节点上执行 **display mpls te ds-te summary** 命令，可查看 DS-TE 的配置信息。以下是在 PE1 上执行该命令的输出示例。

```
[PE1] display mpls te ds-te summary
DS-TE IETF Supported :YES
DS-TE MODE                 :IETF
Bandwidth Constraint Model    :RDM
TEClass Mapping (default):
TE-Class ID      Class Type       Priority
TE-Class 0       0                 0
TE-Class 1       1                 0
TE-Class 2       2                 0
TE-Class 3       3                 0
TE-Class 4       0                 7
TE-Class 5       1                 7
TE-Class 6       2                 7
TE-Class 7       3                 7
```

（9）在 PE 和 P 节点的隧道出方向接口上配置链路带宽。链路总带宽为 500Mbit/s，

BC0 的带宽为 500Mbit/s，BC1 的带宽为 375Mbit/s，BC2 的带宽为 250Mbit/s。注意，RDM 带宽约束模型中，各 CT 间是可共享带宽的，优先级最高的 BC0 具有最高带宽。

PE1 上的配置。

```
[PE1] interface gigabitethernet 3/0/0
[PE1-GigabitEthernet3/0/0] mpls te bandwidth max-reservable-bandwidth 500000
[PE1-GigabitEthernet3/0/0] mpls te bandwidth bc0 500000 bc1 375000 bc2 250000
[PE1-GigabitEthernet3/0/0] quit
```

P 上的配置。

```
[P] interface gigabitethernet 1/0/0
[P-GigabitEthernet1/0/0] mpls te bandwidth max-reservable-bandwidth 500000
[P-GigabitEthernet1/0/0] mpls te bandwidth bc0 500000 bc1 375000 bc2 250000
[P-GigabitEthernet1/0/0] quit
[P] interface gigabitethernet 2/0/0
[P-GigabitEthernet2/0/0] mpls te bandwidth max-reservable-bandwidth 500000
[P-GigabitEthernet2/0/0] mpls te bandwidth bc0 500000 bc1 375000 bc2 250000
[P-GigabitEthernet2/0/0] quit
```

PE2 上的配置。

```
[PE2] interface gigabitethernet 3/0/0
[PE2-GigabitEthernet3/0/0] mpls te bandwidth max-reservable-bandwidth 500000
[PE2-GigabitEthernet3/0/0] mpls te bandwidth bc0 500000 bc1 375000 bc2 250000
[PE2-GigabitEthernet3/0/0] quit
```

以上配置完成后，在 PE 上执行 **display mpls te link-administration bandwidth-allocation interface** 命令，可查看接口的 BC 带宽分配情况。以下是在 PE1 上执行该命令的输出示例，从中可以看出，CT0、CT1 和 CT2 上有可用带宽了。

```
[PE1] display mpls te link-administration bandwidth-allocation interface gigabitethernet 3/0/0
  Link ID:   GigabitEthernet3/0/0
  Bandwidth Constraint Model     :  Russian Dolls Model (RDM)
  Physical Link Bandwidth(kbits/sec)              :  1000000
  Maximum Link Reservable Bandwidth(kbits/sec):  500000
  Reservable Bandwidth BC0(kbits/sec)             :  500000
  Reservable Bandwidth BC1(kbits/sec)             :  375000
  Reservable Bandwidth BC2(kbits/sec)             :  250000
  Reservable Bandwidth BC3(kbits/sec)             :  0
  Reservable Bandwidth BC4(kbits/sec)             :  0
  Reservable Bandwidth BC5(kbits/sec)             :  0
  Reservable Bandwidth BC6(kbits/sec)             :  0
  Reservable Bandwidth BC7(kbits/sec)             :  0
  Downstream Bandwidth (kbits/sec)                :  0
  IPUpdown Link Status                              :  Up
  PhysicalUpdown Link Status                        :  Up
  GracefulUpdown Link Status                        :  DOWN
  ------------------------------------------------------------------------------
  TE-CLASS   CT    PRIORITY   BW RESERVED   BW AVAILABLE   DOWNSTREAM
                               (kbit/sec)     (kbit/sec)    RSVPLSPNODE COUNT
  ------------------------------------------------------------------------------
     0       0        0          0            500000           0
     1       1        0          0            375000           0
     2       2        0          0            250000           0
     3       3        0          0            0                0
     4       0        7          0            500000           0
     5       1        7          0            375000           0
     6       2        7          0            250000           0
```

```
    7     3     7          0          0          0
    8     -     -          -          -          -
    9     -     -          -          -          -
    10    -     -          -          -          -
    11    -     -          -          -          -
    12    -     -          -          -          -
    13    -     -          -          -          -
    14    -     -          -          -          -
    15    -     -          -          -          -
    ---------------------------------------------------------------------------
```

（10）在 PE1 和 PE2 上配置 TE-Class 映射表，分别对应 CT0、CT1 和 CT2，抢占优先级均为缺省的最高值 0。其实可不进行本步配置，直接采用缺省 TE-Class 映射表。

PE1 上的配置。

```
[PE1] te-class-mapping
[PE1-te-class-mapping] te-class0 class-type ct0 priority 0 description For-BE
[PE1-te-class-mapping] te-class1 class-type ct1 priority 0 description For-AF1
[PE1-te-class-mapping] te-class2 class-type ct2 priority 0 description For-AF2
[PE1-te-class-mapping] quit
```

PE2 上的配置。

```
[PE2] te-class-mapping
[PE2-te-class-mapping] te-class0 class-type ct0 priority 0 description For-BE
[PE2-te-class-mapping] te-class1 class-type ct1 priority 0 description For-AF1
[PE2-te-class-mapping] te-class2 class-type ct2 priority 0 description For-AF2
[PE2-te-class-mapping] quit
```

以上配置完成后，在 PE 上执行 **display mpls te ds-te te-class-mapping** 命令，可查看 TE-Class 映射表的信息。以下是在 PE1 上执行该命令的输出示例。

```
[PE1] display mpls te ds-te te-class-mapping
  TE-Class ID        Class Type      Priority        Description
  TE-Class0          0               0                 For-BE
  TE-Class1          1               0                 For-AF1
  TE-Class2          2               0                 For-AF2
  TE-Class3          -               -               -
  TE-Class4          -               -               -
  TE-Class5          -               -               -
  TE-Class6          -               -               -
  TE-Class7          -               -               -
```

（11）在 PE1 和 PE2 上配置显式路径，通过依次指定下一跳的 IP 地址使隧道有固定的报文转发路径。此处配置的显式路径已在第（4）步配置 TE 隧道时调用。

PE1 上的配置。

```
[PE1] explicit-path path1
[PE1-explicit-path-path1] next hop 172.1.1.2
[PE1-explicit-path-path1] next hop 172.2.1.2
[PE1-explicit-path-path1] next hop 3.3.3.9
[PE1-explicit-path-path1] quit
```

PE2 上的配置。

```
[PE2] explicit-path path1
[PE2-explicit-path-path1] next hop 172.2.1.1
[PE2-explicit-path-path1] next hop 172.1.1.1
[PE2-explicit-path-path1] next hop 1.1.1.9
[PE2-explicit-path-path1] quit
```

完成此步骤后，在 PE 上执行 **display explicit-path** 命令，可查看显式路径信息。以下是在 PE1 上执行该命令的输出示例。

```
[PE1] display explicit-path path1
Path Name : path1        Path Status : Enabled
 1        172.1.1.2        Strict        Include
 2        172.2.1.2        Strict        Include
 3        3.3.3.9          Strict        Include
```

（12）在 PE 节点上配置入接口信任的报文优先级以及优先级映射。在 PE1 和 PE2 节点上信任 DSCP 优先级，而在 P 节点上信任 EXP 优先级。由于缺省的 DSCP-LP 和 EXP-LP 映射关系满足本例需求，故不需要进行修改。参见 9.1.6 节表 9-3。

PE1 上的配置。

```
[PE1] interface gigabitethernet 1/0/0
[PE1-GigabitEthernet1/0/0] trust dscp
[PE1-GigabitEthernet1/0/0] quit
[PE1] interface gigabitethernet 2/0/0
[PE1-GigabitEthernet2/0/0] trust dscp
[PE1-GigabitEthernet2/0/0] quit
[PE1] interface gigabitethernet 3/0/0
[PE1-GigabitEthernet3/0/0] trust dscp
[PE1-GigabitEthernet3/0/0] quit
```

P 上的配置。

```
[P] interface gigabitethernet 1/0/0
[P-GigabitEthernet1/0/0] trust exp
[P-GigabitEthernet1/0/0] quit
[P] interface gigabitethernet 2/0/0
[P-GigabitEthernet2/0/0] trust exp
[P-GigabitEthernet2/0/0] quit
```

PE2 上的配置。

```
[PE2] interface gigabitethernet 1/0/0
[PE2-GigabitEthernet1/0/0] trust dscp
[PE2-GigabitEthernet1/0/0] quit
[PE2] interface gigabitethernet 2/0/0
[PE2-GigabitEthernet2/0/0] trust dscp
[PE2-GigabitEthernet2/0/0] quit
[PE2] interface gigabitethernet 3/0/0
[PE2-GigabitEthernet3/0/0] trust dscp
[PE2-GigabitEthernet3/0/0] quit
```

在 PE 上执行 **display qos map-table dscp-lp** 命令，可查看 DSCP 到本地优先级（LP）的映射关系。以下是在 PE1 上执行该命令的输出示例，满足本示例的优先级映射需求。

```
[PE1] display qos map-table dscp-lp
Input DSCP        LP
-------------------
 0                0
...
10                1
...
20                2
...
54                6
55                6
56                7
```

```
57                  7
58                  7
59                  7
60                  7
61                  7
62                  7
63                  7
```

在 PE 上执行 **display qos map-table exp-lp** 命令，可查看 EXP 到本地优先级（LP）的映射关系。以下是在 PE1 上执行该命令的输出示例，满足本示例的优先级映射需求。

```
[PE1] display qos map-table exp-lp
Input EXP       LP
-------------------
0               0
1               1
2               2
3               3
4               4
5               5
6               6
7               7
```

（13）配置 CT 业务的调度方式。把队列 0（对应 CT0、AF1 类型流量）、队列 1（对应 CT1、AF2 类型流量）配置为 WFQ 调度方式，其他队列（包括 CT2、BE 类型流量对应的队列 2）配置为 PQ 调度方式。因为 WFQ 调度方式的时延抖动要远小于 PQ 调度方式，可以更好地满足 AF1 和 AF2 类型业务传输需求。

#　PE1 上的配置。

```
[PE1] qos queue-profile queue-profile1
[PE1-qos-queue-profile-queue-profile1] schedule wfq 0 to 1 pq 2 to 7
[PE1-qos-queue-profile-queue-profile1] quit
[PE1] interface gigabitethernet 3/0/0
[PE1-GigabitEthernet3/0/0] qos te queue-profile queue-profile1
[PE1-GigabitEthernet3/0/0] quit
```

#　P 上的配置。

```
[P] qos queue-profile queue-profile1
[P-qos-queue-profile-queue-profile1] schedule wfq 0 to 1 pq 2 to 7
[P-qos-queue-profile-queue-profile1] quit
[P] interface gigabitethernet 1/0/0
[P-GigabitEthernet1/0/0] qos te queue-profile queue-profile1
[P-GigabitEthernet1/0/0] quit
[P] interface gigabitethernet 2/0/0
[P-GigabitEthernet2/0/0] qos te queue-profile queue-profile1
[P-GigabitEthernet2/0/0] quit
```

#　PE2 上的配置。

```
[PE2] qos queue-profile queue-profile1
[PE2-qos-queue-profile-queue-profile1] schedule wfq 0 to 1 pq 2 to 7
[PE2-qos-queue-profile-queue-profile1] quit
[PE2] interface gigabitethernet 3/0/0
[PE2-GigabitEthernet3/0/0] qos te queue-profile queue-profile1
[PE2-GigabitEthernet3/0/0] quit
```

以上配置完成后，在 PE 上执行 **display qos queue-profile** 命令，可查看已配置的队列模板信息，以下是在 PE1 上执行该命令的输出示例。

```
[PE1] display qos queue-profile queue-profile1
Queue-profile: queue-profile1
Queue   Schedule   Weight   Length(Bytes/Packets) GTS(CIR/CBS)
-----------------------------------------------------------------------------------
0       WFQ        10            -/-            -/-
1       WFQ        10            -/-            -/-
2       PQ         -             -/-            -/-
3       PQ         -             -/-            -/-
4       PQ         -             -/-            -/-
5       PQ         -             -/-            -/-
6       PQ         -             -/-            -/-
7       PQ         -             -/-            -/-
```

（14）在 TE 隧道入节点 PE1 和 PE2 上配置转发邻接，将 TE 隧道发布给邻居节点，使 TE 隧道参与全局的路由计算，其他节点也能使用此隧道。

PE1 上的配置。

```
[PE1] interface tunnel 0/0/1
[PE1-Tunnel0/0/1] mpls te igp metric absolute 1   #---配置 MPLS TE 的度量为指定的度量值 1
[PE1-Tunnel0/0/1] mpls te igp advertise   #---使能转发邻接将 MPLS TE 隧道作为虚拟链路发布到 IGP 网络的功能
[PE1-Tunnel0/0/1] mpls te commit
[PE1-Tunnel0/0/1] mpls
[PE1-Tunnel0/0/1] quit
[PE1] ospf 1
[PE1-ospf-1] enable traffic-adjustment advertise   #---使能 OSPF 1 进程的转发邻接功能
[PE1-ospf-1] quit
```

PE2 上的配置。

```
[PE2] interface tunnel 0/0/1
[PE2-Tunnel0/0/1] mpls te igp metric absolute 1
[PE2-Tunnel0/0/1] mpls te igp advertise
[PE2-Tunnel0/0/1] mpls te commit
[PE2-Tunnel0/0/1] mpls
[PE2-Tunnel0/0/1] quit
[PE2] ospf 1
[PE2-ospf-1] enable traffic-adjustment advertise
[PE2-ospf-1] quit
```

以上配置完成后，在 PE1 或 PE2 上使用 **display ip routing-table** 显示路由信息，通过自动路由中的邻接转发功能，使 PE1 到 5.5.5.9（即 PE4）的出接口选择了 Tunnel0/0/1；PE2 到 4.4.4.9（即 PE3）的出接口选择了 Tunnel0/0/1。

本书是由华为技术有限公司组织编写的工具图书，也是华为ICT认证系列培训教材。本书以华为S系列交换机、AR G3系列路由器为主线，全面介绍了与MPLS隧道技术自身相关的基础知识、静态LSP、LDP LSP、MPLS TE和MPLS-TE隧道建立，以及与MPLS QoS相关的技术原理及配置与管理方法。

本书集系统性、专业性和实用性于一体，既有全面深入的各种技术实现原理的剖析，又有以Step-by-Step方式的详尽配置步骤的介绍，条理清晰，繁而不杂；并且，本书通过大量典型功能应用配置示例，对各种功能配置任务或配置思路进行深入分析，使理论和实践完美结合，学以致用，化繁为简。







ISBN 978-7-115-45648-9
定价：95.00元

分类建议：通信网络技术
人民邮电出版社网址：www.ptpress.com.cn